

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:39:25 ; Search time 37.13 Seconds

(without alignments)
1932.499 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363
Sequence: 1 MGLPRLVCAFLIAACCCPR.....SSGDKRAPDQGEKYIDLRH 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.032802:*
1: /SID5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3341	99.3	646	20	AAV42404
2	3327	98.9	646	20	AAV42405
3	3327	98.9	646	22	AAAB90798
4	647.5	19.3	628	22	ABB50259
5	647.5	19.3	628	22	ABB50290
6	602	18.3	136	21	AAAG00308
7	602	17.9	595	22	ABR11999
8	498.5	14.8	574	21	AAV93969
9	494.5	14.7	582	22	AAAB76858
10	494.5	14.7	583	17	AAAR7230
11	492.5	14.6	583	17	AAW06891

12	492.5	14.6	583	19	AAW47088	Activated leukocyte
13	466	13.9	570	17	AAV97231	Stem cell marker H
14	414.5	12.3	394	21	AAAB3963	Human cancer assoc
15	280.5	8.3	749	22	ABR10297	Human cDNA seq ID
16	277.5	8.3	712	22	AAE07070	Human gene 20 enco
17	269.5	8.0	550	22	ABR10486	Human cDNA seq ID
18	269.5	8.0	550	22	AAU17986	Human immunoglobul
19	266	7.9	3117	21	AAV53667	Sequence g1/332818
20	264.5	7.9	534	21	AAV95323	Pig VCAM, Sus scr
21	253	7.5	269	22	ABR10346	Human cDNA seq ID
22	252	7.5	1380	20	AAV08402	Drosophila sp. ROB
23	252	7.3	1381	20	AAV13564	Drosophila Robo 2
24	246.5	7.3	600	22	AAU29316	Human PRO polypept
25	244.5	7.3	647	14	AAV39686	VCAM-6D/ICAM-1.
26	244.5	7.3	823	22	ABR61552	Drosophila melanog
27	237.5	7.1	736	14	AAV39682	VCAM/ICAM-2. Synt
28	236	7.0	1240	22	ABG03933	Novel human diagno
29	236	7.0	4412	21	AAV53666	Sequence g1/101742
30	235.5	7.0	4393	22	ABR31889	Amino acid sequenc
31	235.5	7.0	4436	22	ABG23285	Novel human diagno
32	234	7.0	848	21	AAV88565	Human NCAM 140KD
33	233	6.9	647	11	AAV08117	Human NCAM 140KD
34	233	6.9	647	14	AAV38549	Vascular cell adhe
35	233	6.9	647	21	AAV59498	VCAM-6D. Synthe
36	233	6.9	1496	20	AAV81030	Human VCAM-1 prote
37	233	6.9	1496	21	AAV70469	Melanoma associate
38	233	6.9	1498	22	ABR11587	Human p53 target m
39	231.5	6.9	753	20	AAV83927	Human peroxidasin
40	231.5	6.9	828	12	AAV13905	Human T85 protein.
41	230.5	6.9	968	22	ABR60068	Rabbit ATRERO-ELAM
42	230	6.8	735	14	AAV38550	Drosophila melanog
43	230	6.8	739	11	AAV08118	VCAM/ICAM-1. Synt
44	230	6.8	739	12	AAV10316	Vascular cell adhe
45	230	6.8	739	13	AAV21081	ICAM-1/2G7 stiaolog

ALIGNMENTS

RESULT	1
AAV42404	AAV42404 standard; Protein; 646 AA.
XX	
AC	AAV42404;
XX	
DT	01-DEC-1999 (first entry)
XX	
DE	Human MUC18 amino acid sequence.
XX	
KW	prostate cancer; melanoma; cell adhesion; glycoprotein;
KW	metastasis; treatment; detection; diagnostic test.
OS	Homo sapiens.
XX	
PN	W09945392-A1.
XX	
PD	10-SEP-1999.
XX	
PF	02-MAR-1999; 99WO-US04850.
XX	
PR	03-MAR-1998; 98US-0076664.
XX	
PA	(UYEM-) UNIV EMORY.
XX	
PI	WU G;
XX	
DR	WPI; 1999-540899/45.
XX	
PT	N-PSDB; AA20930.
XX	
PS	Detection of metastatic prostate cancer, by detection of MUC18 expression in prostate cancer cells -
PS	Claim 17; Page 7; 80pp; English.

XX This is the amino acid sequence of the Human MUC18 cell adhesion
CC glycoprotein, which is expressed on the surface of melanoma cells,
CC and can be used as a marker for prostate cancer.
CC This MUC18 (hMUC18) cDNA sequence is different from the hMUC18 cDNA
CC given in Genbank accession number AAN28882, AAZ20931. The deduced amino
CC acid residues from this cDNA are very different from that published by
CC Johnson's group in 1989.
CC The presence of this glycoprotein has been correlated with the ability
CC of melanomas to metastasize. MUC18 is also associated with normal
CC vascular tissue, and on the smooth muscle of venules, and it expresses
CC sporadically on capillary epithelium.
CC The method can be used as a diagnostic test for prostate cancer which
CC has a relatively high potential for metastasis or which has
CC metastasized. The physician can then choose the appropriate surgical,
CC chemotherapeutic or radiation treatment regimens. In addition, antibody
CC specific to MUC18 can be used to prevent metastasis of PCs.

XX
SQ Sequence 646 AA;

Query Match 99.3%; Score 3341; DB 20; Length 646;
Best Local Similarity 99.5%; Pred. No. 6.8e-227;
Matches 643; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGLPRIVCAFLLAACCCPRVAGVGEABOPAPELVEVEVSTALKGSLQSOGNLSHV 60
DB 1 mglprivcafllaaaccprvagvpgeapelpelveevgstallkqslsgqgnlshv 60
QY 61 DMFSVHKERTLLIFRVROGOGSEPEYEDORSLODRGATLALTOVTPQDERIFLCQGR 120
DB 61 dmfsvhkerllifrvrvgogsepeyedorsldrgatlatltvtpqderiflcqgr 120
QY 121 PRSOEYRIOLRYKAPKEPNIOVNPGLIPVNSKEPEEVATCGRNQYPIPOVIWYKNGRP 180
DB 121 prsgeyriqlrykapeepniqvnplgipvnskepeevatcgrnqypipqvlywkngrp 180
QY 121 prsgeyriqlrykapeepniqvnplgipvnskepeevatcgrnqypipqvlywkngrp 180
QY 181 LKEEKNRVHIQSSQYVESSGLTYLTQSLTKAQLVKEDKDAQFCYCELYNRLPSGNHMKESRE 240
DB 181 lkeeknrhvhiqssqyveessgltlyltqsltkaqlyvkeddaqfycelynrlpsgnhmkesre 240
QY 241 VTPVVFPTKRWLEVPVGMLEKGRVETRCIADNPPRHFSISQNSQSTRAEBETIN 300
DB 241 vtpvvfptkrwlevpvgmlekgrevetrciadnpprhfsisqnsqstreaebetin 300
QY 301 DNGVLVEPARKHSGRVECOGLDLDTMISLSEPOELAVNYSDVRSFAPAREQEGSS 360
DB 301 dngvlvleparkhsgrveecogldldtmislsepoellavnysdvrsfapareqegss 360
QY 361 LTTTCEAESSQDLEFQWLREETGOVLERGPVLDLHDKREAGGGRVAVSPSIPGLNRT 420
DB 361 ltttceassqdlefqwlreetgovlergpvldldhdkreagggrvavspspipglnrt 420
QY 421 QLVNVAIFGPPMAFKRKRWVKNWVNLSCFASGHPRTTSMNNVNGTASQEDDOPORV 480
DB 421 qlvnvaifgppmafkrrkwvknwvnlscfashprrttismnnvngtaseqqedoprv 480
QY 481 LSTLNVLTPTPELLETGVECTASNDIGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
DB 481 lstlnvltptpelletgvectasndigkntsileflvnlsttltpdsnttglstasph 540
QY 541 TRANSTSTERKLPPEESRGVYIAVYICILAVLANVLYFLKKGLPFRSGKQETIL 600
DB 541 tranststerklppeesrgvyiavycilavlanvlyflkkglpfrsgkqetilt 600
QY 601 PPSKSELVEVVKSDKLPEEMGLLOGSSGDKRAPGDOGEYIDLRH 646
DB 601 ppsksellevvksdklpeemgllogssgdkrapgdgqekyidlrh 646

RESULT 2
AAV42405
ID AAV42405 standard; Protein; 646 AA.

XX
AC AAV42405;
XX
DT 01-DEC-1999 (first entry)
XX
DE Human MUC18 amino acid sequence, (Johnson 1989).
XX
KW prostate cancer; melanoma; cell adhesion; glycoprotein;
KW metastasis; treatment; detection; diagnostic test.
XX
OS Homo sapiens.
XX
PN M09945392-A1.
XX
PD 10-SEP-1999.
XX
PF 02-MAR-1999; 99MO-US04850.
XX
PR 03-MAR-1998; 98US-0076664.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Wu G;
XX
XX WPI: 1999-540899/45.
DR N-PSDB; AAZ20931.
XX
PT Detection of metastatic prostate cancer, by detection of MUC18
PT expression in prostate cancer cells -
XX
PS Disclosure: Page 8; 80pp; English.
XX
CC This is the amino acid sequence of the Human MUC18 cell adhesion
CC glycoprotein, which is expressed on the surface of melanoma cells,
CC and can be used as a marker for prostate cancer.
CC This sequence was originally published in 1989 by Johnson et al
CC (Genbank AAN28882).
CC This sequence differs from the one proposed by the inventors, this amino
CC acid sequence is 603 amino acid residues while the present sequence
CC AAZ20930 has 646 amino acid residues.
CC The presence of this glycoprotein has been correlated with the ability
CC of melanomas to metastasize. MUC18 is also associated with normal
CC vascular tissue, and on the smooth muscle of venules, and it expresses
CC sporadically on capillary epithelium.
CC The method can be used as a diagnostic test for prostate cancer which
CC has a relatively high potential for metastasis or which has
CC metastasized. The physician can then choose the appropriate surgical,
CC chemotherapeutic or radiation treatment regimens. In addition, antibody
CC specific to MUC18 can be used to prevent metastasis of PCs.

XX
SQ Sequence 646 AA;

Query Match 98.9%; Score 3327; DB 20; Length 646;
Best Local Similarity 98.9%; Pred. No. 6.5e-226;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGLPRIVCAFLLAACCCPRVAGVGEABOPAPELVEVEVSTALKGSLQSOGNLSHV 60
DB 1 mglprivcafllaaaccprvagvpgeapelpelveevgstallkqslsgqgnlshv 60
QY 61 DMFSVHKERTLLIFRVROGOGSEPEYEDORSLODRGATLALTOVTPQDERIFLCQGR 120
DB 61 dmfsvhkerllifrvrvgogsepeyedorsldrgatlatltvtpqderiflcqgr 120
QY 121 PRSOEYRIOLRYKAPKEPNIOVNPGLIPVNSKEPEEVATCGRNQYPIPOVIWYKNGRP 180
DB 121 prsgeyriqlrykapeepniqvnplgipvnskepeevatcgrnqypipqvlywkngrp 180
QY 181 LKEEKNRVHIQSSQYVESSGLTYLTQSLTKAQLVKEDKDAQFCYCELYNRLPSGNHMKESRE 240
DB 181 lkeeknrhvhiqssqyveessgltlyltqsltkaqlyvkeddaqfycelynrlpsgnhmkesre 240

QY 241 VTVPVPTFEKVMLEVEPYGMLKEGDRVEIRCLADGNPPHFSISKONPSTREAEEETT 300
Db 241 vtvpvptftekvmlevepygmlekgdrveircldgnpphfsisknpstreaeeetln 300
QY 301 DNGVLVLEPARKHSGRTECOGLDIDTMTISLSEPOELLVNVYSDVRVSPAAPERQEGSS 360
Db 301 dngvlvleparkhsgryecqawldtmtislsepellvnvysdvrvspaaperqegss 360
QY 361 LTLTCEAESSODLEFOWMLREETGOVLERGVPVQLHDLRKREAGGRCVAVSPISPGINRT 420
Db 361 ltlceaeessodlefgwlreetdqvlergvpvlqldlrkreaggrycvasvpsipglnrt 420
QY 421 QLVNVAIFGPPWMAFKERKVVWKENMVLNLSCASGHPRTISWVNGTASEODODOPORV 480
Db 421 qlvnlafgppwmafkervvkenmvlnlscasghprtlswvngtaseododoprv 480
QY 481 LSTLNVLTPELLETGVECTASNDLGKMTSIFLELVNLTTLTPDSNTTGTSTASPH 540
Db 481 lstlnvltpelettegvectasndlgkmtsiflevlnlttltpdsnttgtstasph 540
QY 541 TRANSTSTERKLPEPESRGVIVAVICILVAVLGAVALYFLYKKGKLPCCRSGKQETTL 600
Db 541 tranststerklpepesrgvivavivciltvavlgavlyflykkgklpcrrsgkqetl 600
QY 601 PPSRKSELVVEVKSDDLPEEMGLLOGSSGDKRAPDQGEKTYDLRH 646
Db 601 ppsrktselvvevksdklpeemgllqssgdkrapdqqektydlrlh 646
RESULT 3
ID AAB90798
AC AAB90798 standard; Protein; 646 AA.
XX
DT 15-JUN-2001 (first entry)
XX
DE Human shear stress-response protein SEQ ID NO: 96.
XX
KW Human; shear stress-response protein; vascular disease;
XX
OS Homo sapiens.
XX
PN MO200125427-A1.
PD 12-APR-2001.
XX
PE 02-OCT-2000; 2000MO-JP06840.
XX
PR 01-OCT-1999; 99JP-0280976.
XX
PA (KYOM) KYOMA HAKKO KOGYO KK.
XX
PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
XX
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
XX
DR WPI: 2001-266308/27.
XX
DR N-PSDB: AAH02921.
XX
PT DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
XX
XX arteriosclerosis -
XX
XX Claim 60; Page 497-500; 678pp; Japanese.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human shear stress response proteins. These are useful in the
XX diagnosis, treatment and screening of vascular diseases caused by
XX arteriosclerosis, including heart failure, post-PTCA restenosis and
XX hypertension.

SQ Sequence 646 AA;

Query Match 98.9%; Score 3327; DB 22; Length 646;
Best Local Similarity 98.9%; Pred. No. 6.5e-226;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGLPRLVCAFLLAACCCCBVAVGPDEAPAPAEIYVEVSTALTKGSLQSOGNLSHV 60
Db 1 mglprlvcaflaaccpcbavgpdeapapaeiyevegstallkcgslsqsgnlsHV 60
QY 61 DMFSYHKERKTLIFRRQOGOSSEPEYERLSDRGATLALQTVPDERRFLCOGKR 120
Db 61 dmfsyhkerktlifrrqogossepeyerlsldrgatlalqvtpperflcogkr 120
QY 121 PRSGEYRIQLRVYKAPPEERNIOVNPGLIPVNSKEPEEVATCVGRNGPPIOVYKNGRP 180
Db 121 prsgeyriqlrvykapeernigvnpplgvnskepeevatcvgrngppiovykngrp 180
QY 181 LKEENRVHIOSQVVESSGLTTLQSLKAQLYKEKDKDAOFYCELNRYLPSGNHMKESRE 240
Db 181 lkeeknrvhiosqvveessglttlqslkaqlvkekdkdaofycelnrylpsgnhmkesre 240
QY 241 VTVPVPTFEKVMLEVEPYGMLKEGDRVEIRCLADGNPPHFSISKONPSTREAEEETT 300
Db 241 vtvpvptftekvmlevepygmlekgdrveircldgnpphfsisknpstreaeeetln 300
QY 301 DNGVLVLEPARKHSGRTECOGLDIDTMTISLSEPOELLVNVYSDVRVSPAAPERQEGSS 360
Db 301 dngvlvleparkhsgryecqawldtmtislsepellvnvysdvrvspaaperqegss 360
QY 361 LTLTCEAESSODLEFOWMLREETGOVLERGVPVQLHDLRKREAGGRCVAVSPISPGINRT 420
Db 361 ltlceaeessodlefgwlreetdqvlergvpvlqldlrkreaggrycvasvpsipglnrt 420
QY 421 QLVNVAIFGPPWMAFKERKVVWKENMVLNLSCASGHPRTISWVNGTASEODODOPORV 480
Db 421 qlvnlafgppwmafkervvkenmvlnlscasghprtlswvngtaseododoprv 480
QY 481 LSTLNVLTPELLETGVECTASNDLGKMTSIFLELVNLTTLTPDSNTTGTSTASPH 540
Db 481 lstlnvltpelettegvectasndlgkmtsiflevlnlttltpdsnttgtstasph 540
QY 541 TRANSTSTERKLPEPESRGVIVAVICILVAVLGAVALYFLYKKGKLPCCRSGKQETTL 600
Db 541 tranststerklpepesrgvivavivciltvavlgavlyflykkgklpcrrsgkqetl 600
QY 601 PPSRKSELVVEVKSDDLPEEMGLLOGSSGDKRAPDQGEKTYDLRH 646
Db 601 ppsrktselvvevksdklpeemgllqssgdkrapdqqektydlrlh 646
RESULT 4
ID ABB50259
AC ABB50259 standard; Protein; 628 AA.
XX
DT 08-FEB-2002 (first entry)
XX
DE Lutheran blood group protein (BCAM) ovarian tumour marker protein, #6.
XX
XX Ovarian tumour marker gene; human; overexpression; upregulation;
XX epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
XX identification; serous cystadenoma; borderline serous tumour;
XX serous cystadenocarcinoma; mucinous cystadenocarcinoma;
XX mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
XX undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
XX adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
XX immune response pathway; cell proliferation regulation; protein folding;
XX membrane localised; secreted; therapeutic target; cytostatic;
XX gene therapy; vaccine.

OS Homo sapiens.
 XX MO200175177-A2.
 XX 11-OCT-2001.
 XX 03-APR-2001; 2001WO-US10947.
 XX 03-APR-2000; 2000US-194336P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 DR WPI; 2001-626450/72.
 DR N-PSDB; ABA83083.
 XX
 PT Detecting and identifying ovarian tumor, identifying increased risk for
 PT developing ovarian cancer, and determining effectiveness of ovarian
 PT cancer treatment, by measuring expression level of ovarian tumor marker
 PT gene
 XX
 PS Claim 22: Page 55-56; 140pp; English.

XX The invention relates to methods for diagnosing and prognosing ovarian
 CC tumors in an individual via the detection and measurement of the
 CC expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180,
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
 CC ABA83181 and ABA83183). The methods of the invention are useful for
 CC detecting an ovarian tumor in a patient, for identifying an individual
 CC at increased risk for developing ovarian cancer, in tests for
 CC assessing the relative severity of ovarian cancer, in tests for
 CC monitoring a patient in remission from ovarian cancer and in tests for
 CC monitoring disease status in a patient being treated for ovarian cancer.
 CC The methods can additionally be used to identify a particular tumor as
 CC being an ovarian tumor (i.e., an epithelial ovarian tumor selected from
 CC serous cystadenoma, borderline serous tumor, serous cystadenocarcinoma,
 CC mucinous cystadenoma, borderline mucinous tumor, mucinous
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
 CC tumor. The ovarian tumor marker genes of the invention were identified
 CC using SAGE (serial analysis of gene expression) and were found to be
 CC overexpressed in a broad variety of ovarian epithelial tumor cells
 CC relative to normal ovarian epithelial cells. The marker genes are
 CC implicated in immune response pathways, in the regulation of cell
 CC proliferation and in protein folding, and many of these are membrane-
 CC localised or secreted. In addition to their use as diagnostic and
 CC prognostic markers, the ovarian tumor marker genes or their encoded
 CC proteins may be used as therapeutic targets for the treatment and
 CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent
 CC proteins encoded by ovarian tumor marker genes of the invention.

XX Sequence 628 AA;

Query Match 19.3%; Score 647.5; DB 22; Length 628;

Best Local Similarity 29.5%; Pred. No. 3.6e-37;

Matches 202; Conservative 103; Mismatches 265; Indels 115; Gaps 27;

QY 2 GUPR-LVCAFLAACCCCRVAGVPEAEOPAPBELVEVEVGGSTALKCGISQSGNLSH- 59
 DB 12 GPRILLIALLAVIIA-----hpdaqeavrlsvpplvevmgkxvlldc---tpgtctdh 62
 QY 60 -VDWF-----SVHKEKRTLIFRVROGQSEPEYRORSLDRGATLALTOV 106
 DB 63 mlewfltdsgarplaseemgseqlvumhdtgrspp---yqldsgqr---lvlaaa 115
 QY 107 TPQDERIFLC--QGKRRPOEVRIQLRVYKAPPEPNIQVPLGIPVSKPEEYATGVR 164
 DB 116 qvgderdyvcvragaaataarlnvfkpaatevspkkytltsvmedsaqelactstr 175
 QY 165 NGVPIQVIMYKNGRPLK--EKKNRVHIQSSQTV-ESSGLYTLQSIILKAOLVKKDKDAQF 221

DB 176 ngnpapklitwyrngqrlvewmpempegymtsrvtvreaqllslstcljrlrkddrdasf 235
 QY 222 YCELNRLPSCGHHMK-ESREEVTPVFFYPRXK--WL--EVEPYGMKEDGRVETRLADG 276
 DB 236 hcaahyslpegthridspctfhlthyprehvfwgspstpegwvteqdtvgllcrqdg 295
 QY 277 NPPHFISKQNPSTREAEETTNDN--GVLVLEPARKHSGRYECOGDLDTMTSL-LS 333
 DB 296 spspeytlfr-----lqdegevlvnllegnltlegvtirgsgfygcrvedydaadvqls 351
 QY 334 EPQELVNVVSDVRVSPAPPERQEG-----SSLTLCCEAASSODLERQWLREETGV 365
 DB 352 kclrlvayldpbls-----egkvlslplnssavmcsvnhglptpalrlwtdst--p 402
 QY 386 LERGPVLOLHDKREAGGRCVAVSPSIPGLNRTOLVNVAIIRGPPMAFERK-----V 440
 DB 403 lqdgpmllssltfdngcyceaslpvrlstqtnftllvgsspelkcaalepkeds 462
 QY 441 WYKENVNLNLSCEASGHPRPTISWVNGTASQEDQDPQ--VLSTINLVTPLELTGV 497
 DB 463 w-regdevcllcsarghpbdklswsqly-gspaeplpgrgwwsssltkvtsalsrdyl 520
 QY 498 ECTASNDLKNISILFLELVNLTLPDSTTTGLSTASPHTRANSTSTERKLPPES 557
 DB 521 sceasnphnknkhvfhfgavs-----pqlsq 546
 QY 558 RGVIYAVIYICILVAVLGLVLYLYKKGKLPORRSQKQETLTPSRKSELVYEVKSDKL 617
 DB 547 agvaavavavsvglilllvavfycvrrkkg-pccrgrfexgappp---gepglshsgsq 602
 QY 618 PEMGGLIQ--SSGDKRAPDQGEK 640
 DB 603 pegtglmgsagsgarqysgfigde 627

RESULT 5

ABB50290 standard; protein; 628 AA.

ABB50290;

08-FEB-2002 (first entry)

Lutheran blood group (BCAM) ovarian tumor marker protein, SEQ ID NO:70.

XX Ovarian tumor marker gene; human; overexpression; upregulation;
 XX epithelial tumor; cancer; diagnosis; prognosis; disease monitoring;
 XX identification; serous cystadenoma; borderline serous tumor;
 XX serous cystadenocarcinoma; mucinous cystadenocarcinoma; endometrioid carcinoma;
 XX mucinous cystadenoma; borderline mucinous tumor; adenofibroma;
 XX undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 XX adenofibroma; Brenner tumor; serial analysis of gene expression; SAGE;
 XX immune response pathway; cell proliferation regulation; protein folding;
 XX membrane localised; secreted; therapeutic target; cytostatic;
 XX gene therapy; vaccine.

XX Homo sapiens.

XX MO200175177-A2.

XX 11-OCT-2001.

XX 03-APR-2001; 2001WO-US10947.

XX 03-APR-2000; 2000US-194336P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;

XX WPI; 2001-626450/72.

XX N-PSDB; ABA83116.

Db 528 sceasn-----phgnkrhvfhfgtvapt-----sq 553

Qy 558 RGVIYAVIVICLVAVIGAVIYELTKGKLPGRSGOE 597

Db 554 agvawavavsvglilvvvfyvcrkyg-pccqrretek 592

RESULT 8

AA93969

ID AA93969 standard; Protein: 574 AA.

AC AA93969;

DT 03-OCT-2000 (first entry)

DE Amino acid sequence of a partial rat antigen Ag2117.

XX Antigen: Ag2117; pancreatic cell; monoclonal antibody; surface antigen.

XX Rattus sp.

OS WO200037503-A1.

PN 29-JUN-2000.

PD 22-DEC-1999; 99WO-US30741.

PF 22-DEC-1998; 98US-0218539.

PR 22-DEC-1998; 98US-0218539.

XX (RAVE-) RAVEN BIOTECHNOLOGIES INC.

PA Mather JP, Bald LN, Roberts PR, Stephan JF;

PI WPI: 2000-452179/39.

XX N-PSDB; AA57480.

DR Immunizing a host mammal to produce population of monoclonal antibodies

XX that bind to antigens of specific cell type comprises introducing

PT viable cells with serum-free surfaces -

XX

PS Example 10; Fig 10; 61pp; English.

XX The present sequence represents rat antigen Ag2117. The antigen is

CC representative of pancreatic cell lines. The antibody that recognises

CC Ag2117 was identified using the method of the invention. The

CC specification describes a method for immunizing a host mammal to produce

CC a population of monoclonal antibodies that bind to antigens

CC representative of a specific cell type that are heterologous to the

CC host mammal. The method comprises introducing into the mammal intact

CC and viable cells with surfaces which are free of serum. The methods

CC are useful for producing populations of antibodies that bind to antigens

CC representative of a specific cell type, specifically antibodies that

CC are tissue-selective, sub-tissue selective or cell-type specific. These

CC antibodies facilitate the identification of novel antigens and the

CC delineation of the combination of surface antigens on a specific cells

CC type. The antibodies or hybridomas produced using the methods have

CC diagnostic and therapeutic applications.

CC

SO Sequence 574 AA:

Query Match 14.88; Score 498.5; DB 21; Length 574;

Best Local Similarity 24.7%; Pred. No. 1e-26;

Matches 151; Conservative 122; Mismatches 265; Indels 73; Gaps 25;

Qy 5 RVNAFLAACCCCPRAAGPGEAEOPAPRLVEVVGSTLLKCGLSQSGNLSHWDWS 64

Db 10 livcellisaavljpgjgw-----tvsaygdltvmpcorldvpq-nlmfkwxy 58

Qy 65 VHKERTLLIFRVROGOGS-----EPGEYEQRLSDRGATLALVOYTPQDERIFLCO-GK 119

Db 59 ekspgspvflafirsttktskvgdyddvpeykarlslse-nyltslnaklsdektfvcmlvt 117

Qy 120 RPRSQEYRIOLRYAKAPEEENIOWNPGLIPVNSKEBEVATCGRNRPDIPOVIWYKNGR 179

Db 118 ednufaeaprlvkrfkqskpel-vnraaf-leteqklkljgcsidsydgntwyrngk 175

Qy 180 PLKEERKRVHIQSSQIYE-SSGLITLQSLIKAGVLVEDKDAQYCYCLNRLRPSGNMKRS 238

Db 176 vlgpvdgevsillfkkelddpqtqlytmstslsykctksdqmpticsvtyygsqgktyis 235

Qy 239 REYVAPVEYFPEKVMLEEV-EPVGMLEKRGDREIRCLADGNPPH---FSSIKONPSTREA 294

Db 236 eqaifdiyppegvltivlppknaikgedniltqclgngnpppeetlmfyjppqgaetirs 295

Qy 295 EEEETNDNGVLVEPARKESGRYECOGDLDITMISLSEQDLVNVYSDVAVSPAP- 353

Db 296 ntyltd-----vratnagdykcalidqrmaast----ltvhyldslnpsgev 342

Qy 354 ERQGSSTLTLCESQDLEFQWLREPEOVLERGVLOLHMLKREAGGCGYCAVASVPS 413

Db 343 tkqigdtlpscltsaarnltvwmkdnlr--lrssp--sfisalhyqdaqnyvcetalg 398

Qy 414 IPGINRTQLVNVAIFGPPMAFKERKVMKENVNLNLSCEASGHPPTISWVNGTAS-- 471

Db 399 veglkkresltlivegkp--qikmtkktpsglskcltichvegfppkaidwtlgsqsvl 456

Qy 472 -EQDQDPQ--RVLSTNLVTPPELLETGVECTASNDLGKNTSLLELVNLTTLTPDSN 527

Db 457 nteespyingryys--kllispeantv-ltctaenqlertvsnlvsaisl----pe-- 507

Qy 528 TTGTGISTASPHRVANSTSPERKLPEESRGVIVIVICLVAVIGAVIYELX-KRG 586

Db 508 -----headdisdent-ekvndaklivgvlvllaaivagvvywlymkks 554

Qy 587 KLPGRSGOE 597

Db 555 ktaeshakkk 565

RESULT 9

AA976858

ID AAB76858 standard; Protein: 582 AA.

XX AAB76858;

XX 12-APR-2001 (first entry)

DE Human lung tumour protein related protein sequence SEQ ID NO:334.

XX Human: lung cancer; lung tumour; lung tumour protein; gene therapy;

XX lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;

XX cytosolic; antisense inhibition.

XX Homo sapiens.

XX WO200100828-A2.

XX 04-JAN-2001.

XX 30-JUN-2000; 2000WO-US18061.

XX 30-JUN-1999; 99US-0346492.

XX 15-OCT-1999; 99US-0419356.

XX 17-DEC-1999; 99US-0466867.

XX 30-DEC-1999; 99US-0476300.

XX 06-MAR-2000; 2000US-0519642.

XX 22-MAR-2000; 2000US-0533077.

XX 10-APR-2000; 2000US-0546259.

XX 27-APR-2000; 2000US-0560406.

XX 05-JUN-2000; 2000US-0589184.

XX (CORI-) CORIXA CORP.

XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;

PI Retter MW, Mannion J;

```
XX  WPI; 2001-071488/08.
XX
XX  Lung tumor-associated proteins and the nucleic acids that encode them,
PT  useful for preventing, diagnosing and treating lung cancer -
XX
XX  Example 1; Page 262-264; 436pp; English.
XX
CC  The present invention describes immunogenic portions of lung tumour-
CC  associated proteins (I) and the nucleic acids (NAs) that encode them.
CC  (I) have cytostatic activity and can be used in gene therapy, antisense
CC  inhibition and in vaccines. The NAs and the lung tumour-associated
CC  proteins they encode may be used in the prevention, treatment and
CC  diagnosis of diseases associated with their inappropriate expression,
CC  especially lung cancers. For example, the NAs may be administered to
CC  treat diseases by rectifying mutations or deletions in a patient's genome
CC  that affect the activity of the protein by expressing inactive proteins
CC  or to supplement the patient's own production of (I). Additionally, the
CC  NAs may be used to produce the lung-tumour associated protein, according
CC  to standard recombinant DNA methodology. Conversely, antisense NA
CC  molecules may be administered to down regulate protein expression by
CC  binding with the cells own genes and preventing their expression. The NA
CC  and complementary sequences may also be used as DNA probes in diagnostic
CC  assays to detect and quantitate the presence of similar NA sequences in
CC  samples, and hence which patients may be in need of treatment for lung
CC  cancer. The (I) may be used as antigens in the production of antibodies
CC  and in assays to identify modulators (agonists and antagonists) of the
CC  expression and activity of the protein. AAF68083 to AAF68878 and
CC  AAB76848 to AAB76878 represent human lung tumour protein related
CC  nucleotide and protein sequences which are used in the exemplification
CC  of the present invention.
XX
XX  Sequence 582 AA:
SQ

Query Match      14.7%; Score 494.5; DB 22; Length 582;
Best Local Similarity 24.8%; Pred. No. 2e-26;
Matches 145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;

OY  41 GSTALLKCGLSQSGNLSHDMFVSHKEKRTLLFRVROGOGS---EPEEYQRLSD 96
DB  35 gdtlllpcrlidvpq-nlmfgkwyekpdsprvfiafrstsksvqydvdpkydrlnlse 93
OY  97 RGTATLALQVTPQDERIFLCO-GKRRPSQRYRLQRYKAPPEPNQVNLGIPVNSKEP 155
DB  94 -nylslsnaarsnadekrtfcmlytvednvtfaaplvtvfkypskpelvskalfi--eteq1 150
OY  156 EEVATCVGRNGYPIPOVIWYKNGRPLKEEKNRVHIOSQTVES-SGLYTLQSLTKAQLVK 214
DB  151 krlgdcisedsydpdgnltwyrngkvlhplegavvllfkkempvvtqlylmtstleykttk 210
OY  215 EDKDAOFTCELNRYRLPSGNHMKESREVTVPVFYPTKEVWLEV-EPVGMLEKGDREIRCL 273
DB  211 adlqmpftcsvtygypsgqkltlshseqavfdlypteqvliqvlppknaikgednltlkc1 270
OY  274 ADGNPPPH---FSISKONSTREAEETNDNGVLELPARKREHGRECOGLDITMIS 330
DB  271 gngnpppeefilylpgqpegirtsntyltd-----vrrnaqdykcsllidkksmla 322
OY  331 LUSEPQELLVNVSDVRVSPAAP-EROEGSSLTLCCEASSODLEFQWLRREFTGOVLERG 389
DB  323 sta-----ltvhy1-dlslnpsegtvtrgdalpvscitssasnatvwmkdlir--lss 375
OY  390 PVLQHLHLKREAGGRCYASVPSIPGLNRITQVLVNAIFGPWMAFKERKRVWVKNMVLN 449
DB  376 p--sfslshygdagnyvetalvegejlkreslclivegkp-qikmtkktdpjslakt 431
OY  450 LSCESGHPRTISNVNGTAS---EQODDPQ---RVSLTNLVLTPLLELGVGVCTASN 503
DB  432 lichenegfipkpaqlvcltligsgsvnlqteespyngrys--klisspeant-ltcten 488
OY  504 DLGKNTSILFELVNLTLTPDSNTTGTLSSTASPHRANSTSERKLPEESGAVTIV 563
DB  530 givvg11laavagvywlymkkstaskhvnkdglgmeenkkle 574
```

```
DB  489 qlertvsnlvsaisl-----pe-----hdeadeisdent-ekvndqakliiv 529
OY  564 AVIYCILVLAVALVAFELY-KKGLPCRRSGKQETILPPSRKSE 607
DB  530 givvg11laavagvywlymkkstaskhvnkdglgmeenkkle 574

RESULT 10
AAR97230
ID  AAR97230 standard; Protein: 583 AA.
XX
XX  AAR97230;
AC  AAR97230;
DT  08-OCT-1996 (first entry)
XX
XX  Stem cell marker HCAPro.1.
XX
XX  Stem cell marker; haematopoietic cell antigen; HCA; HCAsq.1;
KW  HCAPro.1.
XX
XX  Homo sapiens.
XX
XX  EP716146-A2.
XX
XX  12-JUN-1996.
XX
XX  02-FEB-1995; 95EP-0300661.
XX
XX  06-DEC-1994; 94US-0352323.
XX
XX  (SYST-) SYSTEMIX INC.
XX
XX  Gearing D, Uchida N, Yang Z;
XX
XX  WPI: 1996-269990/28.
XX
XX  N-PSDB: AAT28819.
XX
XX  DNA encoding human haematopoietic cell antigen proteins - useful as
XX  stem cell marker proteins in functional studies and for antibody
XX  prodn.
XX
XX  Claim 8; Fig 4A; 17pp; English.
XX
XX  HCAPro.1 (AAR97230) and HCAPro.2 (AAR97231) are novel human
XX  haematopoietic cell antigens associated with stem cells. They
XX  are the respective products of nucleic acid sequences HCAsq.1
XX  (AAT28819) and HCAsq.2 (AAT28820) obtd. from human CD34+ bone marrow.
XX  Expression systems are provided for prodn. of recombinant HCA
XX  proteins. The proteins are useful as stem cell marker proteins in
XX  functional studies, and can also be used to produce antibodies
XX  that allow the purification of stem cells from haematopoietic and
XX  other sources.
XX
XX  Sequence 583 AA:
SQ

Query Match      14.7%; Score 494.5; DB 17; Length 583;
Best Local Similarity 24.8%; Pred. No. 2e-26;
Matches 145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;

OY  41 GSTALLKCGLSQSGNLSHDMFVSHKEKRTLLFRVROGOGS---EPEEYQRLSD 96
DB  36 gdtlllpcrlidvpq-nlmfgkwyekpdsprvfiafrstsksvqydvdpkydrlnlse 94
OY  97 RGTATLALQVTPQDERIFLCO-GKRRPSQRYRLQRYKAPPEPNQVNLGIPVNSKEP 155
DB  95 -nylslsnaarsnadekrtfcmlytvednvtfaaplvtvfkypskpelvskalfi--eteq1 151
OY  156 EEVATCVGRNGYPIPOVIWYKNGRPLKEEKNRVHIOSQTVES-SGLYTLQSLTKAQLVK 214
DB  152 krlgdcisedsydpdgnltwyrngkvlhplegavvllfkkempvvtqlylmtstleykttk 211
OY  215 EDKDAOFTCELNRYRLPSGNHMKESREVTVPVFYPTKEVWLEV-EPVGMLEKGDREIRCL 273
```

```

Db      212 adqmpficsvlyygsqgkltinsegavfdiypfteqvltvlpknaikgdnltlckl 271
QY      274 ADGNPPH-----FSISKONPSTREAEERTNDNVLEPARKHSGREGQGLDTMTS 330
Db      272 gngnppreeflfypgperitrsnytltd-----vrratdykcsllidkksmla 323
QY      331 LLSPEOLLVNVSDVAVSPAAP-EROEGSLTLTCEAESSODLEQWLREETGQYLENG 389
Db      324 sta-----ltvhyt-dlslmpsgvtrtqgdalpvscitssarnatvwmkdnlr--lrs 376
QY      390 PVLQHLKREAGGRCVAVSPISPGLNRTQLVNVAIGPPMAKREKRVWKEMMVLN 449
Db      377 P--sfslhyqdagnyvcetalqevglkresltlvegkp--qikmtkktqpglskt 432
QY      450 LSCASGHPRTISMVNWGTAS---EODDPQ---RVLSLTNLVTPLELTGVECTASN 503
Db      433 lchvegfpkpaqwtltsqsvintgeespyingryys--klispeenvt-ltctaen 489
QY      504 DLGKNTSILFLELVNLTTLTPDSNTTGTSTASPHSTRANSTERKLPPEPSRGVTV 563
Db      490 qlertvnslnvsaisl-----pe-----hdeadeisdenr-ekvndqakliv 530
QY      564 AVIVCIIVLAVLGAVLVLYL-KKGKLPCKRRSGKQETITPPSRKSE 607
Db      531 givvgjlllaalvagvvywlymkkskaskhvnkldlgmneenkle 575

RESULT 11
AAW06891
ID      AAW06891 standard; Protein; 583 AA.
AC
XX
AC      AAW06891;
XX
DT      18-MAR-1997 (first entry)
XX
DE      Human activated leukocyte-cell adhesion molecule ALCAM.
XX
KM      Activated leukocyte-cell adhesion molecule; ALCAM; CD6 ligand;
KW      thymic epithelial cell; inflammation; allograft rejection;
XX      neurodegenerative disease.
XX
OS      Homo sapiens.
XX
FH      Key
FT      Peptide
FT      1..27
FT      /label= Sig_peptide
FT      28..583
FT      /label= Mat_protein
FT      28..527
FT      /label= Extracellular_domain
FT      /note= "the extracellular domain can be subdivid"
FT      528..551
FT      /label= Transmembrane_domain
FT      552..583
FT      /label= Extracellular_domain
FT      91
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      95
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      167
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      258
FT      /label= Asn_Ser
FT      /note= "variation owing to polymorphism in HL60
FT      265
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      Misc-difference
FT      301

```

```

FT      /label= Met_Thr
FT      /note= "variation owing to polymorphism in HL60
FT      306
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      337
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      361
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      457
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      480
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"
FT      499
FT      /label= Glycosylation
FT      /note= "putative N-glycosylation site"

W09634880-A1.
07-NOV-1996.
29-APR-1996; 96WO-US06010.
01-MAY-1995; 95US-0432016.
(BRIM ) BRISTOL-MYERS SQUIBB CO.
(UYDU-) UNIV DUKE.
Aruffo A, Bowen MA, Haynes BF, Margardt H, Patel D;
WPL: 1996-506097/50.
N-PSDB; AAT46075.
CD6 ligands present on surface of thymic epithelial cells - used to
develop prods. for treating e.g. inflammation, organ allograft
rejection or neurodegenerative diseases
Example 4; Fig 29; 112pp; English.
XX
PS      Human activated leukocyte cell adhesion molecule (AAW06891) or ALCAM
XX      is a CD6 ligand present on the surface of thymic epithelial cells,
XX      monocytes, activated T-cells, and other cells. Its amino acid
XX      sequence was deduced from cDNA clones (AAT46075) obt'd. from HL60
XX      and T-cell lines. ALCAM polypeptides can be produced, e.g. as a
XX      fusion proteins, in transformed host cells. They are useful for
XX      inhibiting CD6/CD6 interactions between cells and to screen for
XX      agents that modulate this interaction. Inhibitors can be used to
XX      treat e.g. inflammation, multiple sclerosis, inflammatory uveitis,
XX      rheumatoid arthritis, T-cell mediated vasculitis syndromes, organ
XX      allograft rejection and neurodegenerative diseases.
XX
SQ      Sequence 583 AA;

Query Match 14.6%; Score 492.5; DB 17; Length 583;
Best Local Similarity 24.8%; Pred. No. 2,7e-26;
Matches 145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;

QY      41 GSTALKCGLSQSGNLSHYDWFSEVHKERKTLIFRVQGGGOS---EPGEYQRSLQD 96
Db      36 gdtllipcorldvpg-nlmfigkwkyekdgsprvfiafstsksvqydvdyeydrnlise 94
QY      97 RGAFLALTYQTPQDERIFLQ-QKRRSQEYRIQLRVYKAAEPPNQLVPLGIYVNSKEP 155
Db      95 -nytlisnarisdekrfvcmlvtednvfeapltivkvfkpskpeylvskaflf--eteql 151
QY      156 EEVATCGVRNGYRIPQYIWKNGRPLKEKNRHHIOSQVLES-SGLTYLQSLTKAQLYK 214

```

```

Db      152 kklgdcisedspdgnticwyrngkvlhpllegavvllfkemdpvrcqlytmstleyctk 211
OY      215 EKDQAOFCNELNRLPSGNHMKESREVTVPVPEPEKMLEV-EPVGNLKGDRYEIRCL 273
Db      212 adlqmpftcsvlygypsgqkthseqavfdiypgelyqlvlpkxakagdnltlxl 271
OY      274 AGGNPPH---PSISKONSTRAESEETTNDGNVLLEPARKEHSGRECOGLDIDTMS 330
Db      272 gngnppeefilylpgqpegitssntlylx-----vrrnagdykcsldkkm1a 323
OY      331 LUSEPOELLVNVSDVRVSPAAP-ERQEGSSITLTCEASSODEFQWLREBTQVLERG 389
Db      324 sta-----ltvhy1-dlshnpsevrtrlgdaipvscslasrnaivvmkdn1ir--lss 376
OY      390 PVLQHLDLKREAGGRCVASTPISGLNRQLVNAVAFGPPMMAFKRRKVMVKNEMVLN 449
Db      377 p--sfslhygdagayvcetalqevglkkrslclivegkp--qikmtktqpsgiskt 432
OY      450 LSCASGHPRTISWNVNGTAS---EODDPQ---RVLSTLVNLTPELLETGVECTASN 503
Db      433 l1chvegfpkpalqwtlgsavlnqteespylmgrysa--kllspsenvt-lctaaen 489
OY      504 DLGKNTSILFELVNLTLTPDSNTTGTGLSTASPHRANSTSTERKLPRESRGVYIV 563
Db      490 qlertvnslnvaslsl---pe-----hedeaisdent-ekvndqakl1v 530
OY      564 AVYICVLVAVLCAVLYFLY-KKGKLPCCRSGKOETITPPSRKSE 607
Db      531 glvvgjlllaalyagvyw*lymkksktaahvnhkd1gmmeenkle 575

RESULT 12
AAW47088
ID      AAW47088 standard; Protein; 583 AA.
AC      AAW47088;
DE      20-JUL-1998 (first entry)
XX      Activated leukocyte-cell adhesion molecule (ALCAM).
XX      ALCAM; activated leukocyte-cell adhesion molecule; CD6 ligand;
XX      antibody; human.
XX      Homo sapiens.
FH      Key
FT      1.27 Location/Qualifiers
FT      /label= Sig_peptide
FT      28..583
FT      /label= Mat_protein
FT      28..527
FT      /note= "extracellular domain"
FT      528..551
FT      /note= "transmembrane domain"
FT      552..583
FT      /note= "cytoplasmic domain"
FT      91
FT      /note= "N-glycosylated"
FT      95
FT      /note= "N-glycosylated"
FT      167
FT      /note= "N-glycosylation"
FT      258
FT      /label= Asn_Ser
FT      /note= "encoded by ART (polymorphism in nucleotide
FT      sequence"
FT      Modified-site
FT      265
FT      /note= "N-glycosylated"
FT      301
FT      /label= Met_Thr
FT      /note= "encoded by AVG (polymorphism in nucleotide
FT      sequence)"

```

```

FT      Modified-site 306
FT      /note= "N-glycosylated"
FT      Modified-site 337
FT      /note= "N-glycosylated"
FT      Modified-site 361
FT      /note= "N-glycosylated"
FT      Misc-difference 447
FT      /note= "encoded by CAR"
FT      Modified-site 457
FT      /note= "N-glycosylated"
FT      Modified-site 480
FT      /note= "N-glycosylated"
FT      Modified-site 499
FT      /note= "N-glycosylated"
XX      PN      WO9803551-A1.
XX      PD      29-JAN-1998.
XX      XX      17-JUL-1997; 97WO-US12511.
XX      PF      18-JUL-1996; 96US-0684594.
XX      PR      (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX      PA      (UTDU-) UNIV DUKE.
XX      PI      Aruffo A, Bowen MA, Haynes BF, Margardt H, Patel D;
XX      PI      Sladak AW.
XX      DR      WPI: 1998-120703/11.
XX      DR      N-PSDB: AAV13954.
XX      PS      Claim 3; Fig 29; 130pp: English.
XX      CC      This polypeptide is a human CD6 ligand (see AAW47088), designated
XX      CC      activated leukocyte-cell adhesion molecule (ALCAM), that is present
XX      CC      on the surface of thymic epithelial cells, monocytes, activated
XX      CC      T cells and a variety of other cells. Its amino acid sequence was
XX      CC      deduced from cDNA clones (see AAV13954) isolated from human T cell
XX      CC      and HL60 cDNA libraries. It is the human homologue of chicken
XX      CC      BEN, and shows homology to neuroilin, RAGE and MUC18. Anti-CD6
XX      CC      ligand antibodies, or their binding fragments, are useful for
XX      CC      inhibiting binding of CD6 present on the surface of a first cell
XX      CC      to that present on a second cell. CD6 ligands and anti-CD6 ligand
XX      CC      antibodies are also useful in screening test compounds for the
XX      CC      ability to inhibit binding of CD6 ligand to an anti-CD6 ligand
XX      CC      antibody (all claimed).
XX      SQ      Sequence 583 AA:

Query Match 14.6%; Score 492.5; DB 19; Length 583;
Best Local Similarity 24.8%; Pred. No. 2,7e-26;
Matches 145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;

OY      41 GSTALLKCGLSQSGNLSHVDVFSVHKEKRTLIRVRQGGQS-----EPGIEORLSLD 96
Db      36 gdtllipcrldwpg-nlmgfkwyekqpsvfiatfsstkksqyddvpeykdrlnlse 94
OY      97 KQATLALQVTPQDERITLCO-GKRPRSQEYRIQLRVYKAPEDNPIONPLGIVNSKEP 155
Db      95 -nytlstlnarisdskrtvcmlvtednvfeapcltvkvtqpskpeivskalfi--eteq1 151
OY      156 EVATACVGRNGYPIPIQVITWYKGRPLKEKNRVHIQSSQTVES-SGLYTLQSIKAOLVK 214
Db      152 kklgdcisedspdgnticwyrngkvlhpllegavvllfkemdpvrcqlytmstleyctk 211
OY      215 EKDQAOFCNELNRLPSGNHMKESREVTVPVPEPEKMLEV-EPVGNLKGDRYEIRCL 273

```

Db	212	adlmpficsvlyvypsgqkltihsegafdiypteqvtlqvlpbpkkaiknegdnitlckl	271
Qy	274	ADGNPPPH---FSISKONPSTREAEEETTNDNGVLLEPAKREHSGRYECQGLDLDTMIS	330
Db	272	gnqpppeeflfiypgqpegfirsntyltd-----vrrnatgdykcslldkkemia	323
Qy	331	LISEQELLVNYSVDVRSPPAP-ERQEGSSLTITCEAESQDLEFQWLEEFQGVLERG	389
Db	324	sta---ltvhytl-dlslnpsegrtlqigdalpvcsltsaarnatvwmkdnlt--lrss	376
Qy	390	PVLQIHDIKREKGGYRCVAVSPSPICGLNRIQNVVAIFCPNMAFERKRYKENVNVLN	449
Db	377	p--sfsslhqdaagnyvcetaldgevgjkkresltllvegmp-qikmtkktdpsjst	433
Qy	450	LSCSAAGPRPRTISWNVNGTAS---EEDQDPQ---RVLSITNLVLTPELLETGYECFASN	500
Db	433	llchvegfpkpaiqwtlctgsgvlnqteespynqrrys--klispeenvt-ltctae	483
Qy	504	DLGKMTSLFLELVNLTLLTPDSNNTTGLSTASPHRANSTSTERKLPEPSRGVIV	567
Db	490	qlertvnslnvsaisi---pe-----hedeaisdeem-ekvndqaklly	533
Qy	564	ATVTCILVLAVLAVLYELY-RRGKLPCRRSKGQETLPPSRKSE	607
Db	531	glvvglllaalvagvyvlymksktashkvdldgmneonkkle	575
RESULT	13		
AA97231	ID	AA97231 standard; Protein; 570 AA.	
XX	AA97231;		
XX	08-OCT-1996	(first entry)	
DE	Stem cell marker HCAPro. 2.		
KW	Stem cell marker; haematopoietic cell antigen; HCA; HCASeq.2;		
XX	HCAPro. 2.		
OS	Homo sapiens.		
PN	EP716146-A2.		
XX	12-JUN-1996.		
PF	02-FEB-1995; 95EP-0300661.		
XX	06-DEC-1994; 94US-0352323.		
XX	(SYST-) SYSTEMIX INC.		
XX	Gearing D, Uchida N, Yang Z;		
DR	WPI: 1996-269990/28.		
XX	N-PSDB: AAT28820.		
XX			
XX	Claim 8; Fig 4B; 17pp; English.		
XX	HCAPro.1 (AA97230) and HCAPro.2 (AA97231) are novel human		
CC	haematopoietic cell antigens associated with stem cells. They		
CC	are the respective products of nucleic acid sequences HCASeq.1		
CC	(AAT28819) and HCASeq.2 (AAT28820) obtd. from human CD34+ bone marr		
CC	Expression systems are provided for prodn. of recombinant HCA		
CC	proteins. The proteins are useful as stem cell marker proteins in		
CC	functional studies, and can also be used to produce antibodies		
CC	that allow the purification of stem cells from haematopoietic and		
XX	other sources.		

Seq	Sequenc	570 AA:
QY	Query Match	13.9%; Score 466; DB 17; Length 570;
DB	Best Local Similarity	24.4%; Pred. No. 1.9e-24;
Matches 143;	Conservative 109;	Mismatches 257; Indels 76; Gaps 22.
QY	41 GSTLLKCGISQSGNLSHDMFVHKERKLTIFPVRGOGS---	EPGEQRSLSD 96
DB	36 gdtlliprdlvpq-nlmfgkwkyekpdpsspfiafststksvgyddvpeykarlnlse	94
QY	97 RGATLALVQYPTDERIFLCQ-GKRRSQEYRIQLRYKAEERINQVNPGLIVNSKEP	155
DB	95 -nyltisnariseskrifvcmlvtedvfeapltykvikfqpkspeivskalfi--eteq	151
QY	156 EEVATCGARNQYRIPQVIWYKNGRPKEKRNHYHIOSSQTVES-SGLYTLQSLIAQLVK	214
DB	152 kllgdcisedsyprgnltwngntvlhplegavvllfkfkwemvprqlyltmstleytk	211
QY	215 EDKDAFYCELNLYRLPSGNHMKESREVTYVPEYKWLKV-EPVGMLEKGDVYRCL	273
DB	212 adimprfcsvtylpgpsgqkflhseqvfdlypteqvlyqlpbnkavkegdnltlcl	271
QY	274 ADGNPPH---FSTSKQNPSTRAEETNTDNCVVLLEPARKEHSGRECOGLDITWIS	330
DB	272 gngnpppeefltylpgqpegirtsnytlnd-----vrrnagdykcsalikkmsia	323
QY	331 LTSEQELIVNVSPDVRVSPAAP-ERQEGSSLTLCBASSODLEFQWLRETFQVLERG	389
DB	324 sta-----ltvnyl-dlslnpgevtrtqgdalpvscitiasratvymkdnlr--lrs	376
QY	390 PVLQLHLDKREAGGGRVAVSPRLPDLNTOLVNVAIGCPPMMAKERKVVYKENVYN	449
DB	377 p-sfsishygdaagynvetalgeveglkkrsteltivegk--qlkmktktdpsgjskt	432
QY	450 LSCGASGHPRIPISNVNGTAS---EGDQDPQ---RYLSTLVNLYPELLETGVECTASN	503
DB	433 lchegvfpkpiqvtligsgsvlnqteespylmgrys--ktlispennvt-ltctaa	489
QY	504 DIGKNTSLLELVNLTTLTPDPSNTTGLSLSTASPTFRANSSTERKLPDEESRQVIV	563
DB	490 ql-----etvnslnvsaanesrekvn-----dqaklly	517
QY	564 AVYVCILVAVLGAVLYFLY-KKGLPCRSGKQETILPSPRSKE 607	
DB	518 glavgillalaavgywylmkkstktaskhvnkldgmeentkile 562	
RESULT 14		
ID	ABA43943	standard; Protein: 394 AA.
AC	ABA43943:	
DT	08-FEB-2001	(first entry)
XX	Human cancer associated protein sequence SEQ ID NO:1388.	
XX	Human: cancer associated gene; cancer antigen; detection; cancer;	
KW	diagnosis; cytostatic; proliferative; vlnnerary; immunomodulator;	
KW	antidiabetic; antilasthmatic; antirheumatic; antiarthritic; antiviral;	
KW	antiflammatory; antihypertoid; antiallergic; antibacterial; cardiatic;	
KW	dematological; neuroprotective; thrombolytic; coagulant; neutropic;	
KW	vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;	
KW	immune disorder; haematopoietic cell disorder; autoimmune disorder;	
KW	allergic reaction; graft versus host disease; organ rejection;	
KW	haemostatic; thrombolytic; cardiovascular disorder; infection;	
OS	Homo sapiens.	
XX	WO200053530-A1.	
PN		
XX		


```

Db 297 lgsdvlctcwvgnpplltlwtkkdsnmgprppgsppaalasaqlsnsnqlllkevta 356
QY 401 AGGGRVAVSVPSISPOLNRTQLVNVVAIFGPPMAAFKKRWYKENVVNLSCBASGHRP 460
Db 357 daqlycraivprlgyaere--vplyvngpplisseavgyavrgdgg-kvecfisltpp 413
QY 461 ---TISMNVN---GTAS---EODDOPQVRLSTLNVLPET-LETGVECTASNDLGKN 508
Db 414 dilaawakenflevgfletyvertngsgvstltcinymeadfqlhinctamsfgpg 473
QY 509 TSILFELVNLTTLPDSTNTTGLSTSTASPHTRANSTSTERK-LPEPESRGVIVAVI- 566
Db 474 tailqle-----erevlpv---gliagatig 496
QY 567 VCILVLAVALGAVLYFYKKGKLPGRSGKOEITLPPSRKSELVVE 611
Db 497 asilllfflaivfilyir-----rkgsarkdvcll--rkldikve 533

```

Search completed: June 28, 2002, 10:45:49
 Job time: 384 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:43:45 ; Search time 18.72 Seconds

(without alignments)
842.892 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363
Sequence: 1 MGLPRIVKAFILAAACCCPR.....SSGDKRAPDQGEKTYIDLRH 646Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match length	ID	Description
No.	Score	DB		
1	2302.5	68.5	486 2	US-08-432-016-6 Sequence 6, Appl
2	2302.5	68.5	486 2	US-08-684-594-6 Sequence 6, Appl
3	492.5	14.6	583 2	US-08-432-016-2 Sequence 2, Appl
4	492.5	14.6	583 2	US-08-684-594-2 Sequence 2, Appl
5	444.5	13.2	477 2	US-08-432-016-3 Sequence 3, Appl
6	444.5	13.2	477 2	US-08-684-594-3 Sequence 3, Appl
7	414.5	12.3	466 2	US-08-432-016-4 Sequence 4, Appl
8	414.5	12.3	466 2	US-08-684-594-4 Sequence 4, Appl
9	252	7.5	1381 4	US-09-540-245A-16 Sequence 16, Appl
10	244.5	7.3	647 5	PCT-US93-00031-23 Sequence 23, Appl
11	237.5	7.1	736 5	PCT-US93-00031-15 Sequence 15, Appl
12	233	6.9	647 4	US-09-009-490A-91 Sequence 91, Appl
13	233	6.9	647 4	US-08-482-073-5 Sequence 5, Appl
14	233	6.9	647 4	PCT-US93-00031-11 Sequence 11, Appl
15	232.5	6.9	1260 4	US-08-506-296B-21 Sequence 21, Appl
16	231.5	6.9	828 1	US-08-261-304-2 Sequence 2, Appl
17	230	6.8	735 5	PCT-US93-00031-13 Sequence 13, Appl
18	230	6.8	739 5	US-08-482-073-6 Sequence 6, Appl
19	230	6.8	739 5	PCT-US93-00031-9 Sequence 9, Appl
20	226.5	6.7	662 1	US-08-261-304-7 Sequence 7, Appl
21	226	6.7	738 3	US-08-478-208-32 Sequence 32, Appl
22	225.5	6.7	1253 3	US-08-506-296B-14 Sequence 14, Appl
23	225	6.7	698 2	US-08-602-725-36 Sequence 36, Appl
24	225	6.7	734 2	US-08-389-459A-17 Sequence 17, Appl
25	225	6.7	734 2	US-08-987-867A-17 Sequence 17, Appl
26	224.5	6.6	1651 4	US-09-540-245A-18 Sequence 18, Appl
27	223	6.6	738 6	Patent No. 5264554-2

28	222.5	6.6	644 5	PCT-US93-00031-21 Sequence 21, Appl
29	220.5	6.6	740 5	PCT-US93-00031-17 Sequence 17, Appl
30	220.5	6.6	1241 4	US-09-040-774-2 Sequence 2, Appl
31	219	6.5	642 1	US-08-217-299-1 Sequence 1, Appl
32	218.5	6.5	318 2	US-08-633-148-4 Sequence 4, Appl
33	218.5	6.5	318 2	US-08-633-148-4 Sequence 4, Appl
34	216	6.4	630 2	US-08-752-307B-14 Sequence 14, Appl
35	215	6.4	643 5	PCT-US93-00031-19 Sequence 19, Appl
36	215	6.4	1268 4	US-08-506-296B-28 Sequence 28, Appl
37	213.5	6.3	1356 1	US-08-810-116-8 Sequence 8, Appl
38	213.5	6.3	1356 2	US-07-930-548A-8 Sequence 8, Appl
39	212	6.3	1395 4	US-09-540-245A-15 Sequence 15, Appl
40	207	6.2	788 1	US-08-232-538-15 Sequence 15, Appl
41	207	6.2	788 2	US-08-786-164-15 Sequence 15, Appl
42	205.5	6.1	1356 4	US-09-098-707A-2 Sequence 2, Appl
43	205.5	6.1	1356 4	US-09-483-539-2 Sequence 2, Appl
44	204	6.1	612 2	US-08-752-307B-11 Sequence 11, Appl
45	203.5	6.1	1266 4	US-08-506-296B-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-432-016-6
Sequence 6, Application US/08432016
Patent No. 5968768
GENERAL INFORMATION:
APPLICANT: HAYNES, BARTON F.
APPLICANT: ARUFFO, ALEANDRO
APPLICANT: PATEL, DHAVALKUMAR
APPLICANT: BOWEN, MICHAEL A.
APPLICANT: MAROUARDT, HANS
TITLE OF INVENTION: CD6 LIGAND
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentip Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,016
FILING DATE: 01-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/333,350
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/143,903
FILING DATE: 02-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-08-432-016-6

```

Query Match          68.5%; Score 2302.5; DB 2; Length 486;
Best Local Similarity 91.5%; Pred. No. 2,3e-178;
Matches 450; Conservative 9; Mismatches 22; Indels 11; Gaps

QY      40 VGSYLLKCGLSQSOGNSHVDMSVHKKEKTLFLIRVYGOGGQSEPGYEORLSTODRGA 99
         | | | | | | | | | | | | | | | | : : : | | | | | |
DB       1 VGSYLLKCGLSQSOGNSHVDMSVHKKEKTRSSVCARARANLSTSRISLSDRGA 60
QY      100 TIALTVTPDERITLCOGKR--PR---SQEYRIQLRYAKAPDEBPINOVLGIIPVNSKE 15
        | | | | | | | | | | | | | | | | : : : | | | | | |
DB       61 TIALTVTPDERITLCOGKRGRPRSTASSASTKLRM-----PNIQVNLGIIPVNSKE 11
QY      155 PEEVATCGRNRCYPIPOVIWKNGRPLEEKNRVHIOSQTVESSGLYTLOSILKAOLVK 21
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       115 PEEVATCGRNRCYPIPOVIWKNGRPLEEKNRVHIOSQTVESSGLYTLOSILKAOLVK 17
QY      215 EDKDAOFYCELNLYRLPSGNHHKESREYVPVFYPPEKVMLEVEEPYGMLEKEDRVEIRCLA 27
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       175 EDKDAOFYCELNLYRLPSGNHHKESREYVPVFYPPEKVMLEVEEPYGMLEKEDRVEIRCLA 23
QY      275 DGNPPPHSISKQNPNSTREAEETTNDNGVYLVEPAKKESHSGRRBCOGLDITMISLSE 33
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       235 DGNPPPHSISKQNPNSTREAEETTNDNGVYLVEPAKKESHSGRRBCOAMNIDTMSLSE 29
QY      335 POELLVNVYSIDRVSPARPAREOESSLTLCGEASSODLEFOWIREFTGOVLEERGPIVOL 39
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       235 POELLVNVYSIDRVSPARPAREOESSLTLCGEASSODLEFOWIREFTDOVLEERGPIVOL 35
QY      355 HDLKREAGGGYRCVASVPSIGPLNTLOLVNAIFGPWMAKERKVYKENVMLNLSCEA 45
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       355 HDLKREAGGGYRCVASVPSIGPLNTLOLVNAIFGPWMAKERKVYKENVMLNLSCEA 41
QY      455 SGHPRTISMVNVGASPDODDPQRYLSTLVNLTPELLETGVCTASNDGKMTSLFL 51
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       415 SGHPRTISMVNVGASPDODDPQRYLSTLVNLTPELLETGVCTASNDGKMTSLFL 47
QY      515 ELVNLTTLTLPDS 526
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       475 ELVNLTTLTLPDS 486

RESULT      2
US-08-684-594-6
Sequence 6, Application US/08684594
Patent No. 5998172
GENERAL INFORMATION:
APPLICANT: HAYNES, BARTON F.
APPLICANT: ARUFFO, ALEJANDRO
APPLICANT: PATEL, DHAVALKUMAR
APPLICANT: BOWEN, MICHAEL A.
APPLICANT: MARQUARDT, HANS
TITLE OF INVENTION: CD6 LIGAND
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/684,594
FILING DATE: 18-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,016
FILING DATE: 01-MAY-1995

```

```

1  PRIOR APPLICATION DATA:
2  APPLICATION NUMBER: US 08/333,350
3  FILING DATE: 02-NOV-1994
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER: US 08/143,903
6  FILING DATE: 02-NOV-1993
7  ATTORNEY/AGENT INFORMATION:
8  NAME: WILSON, MARY J.
9  REGISTRATION NUMBER: 32,955
10 REFERENCE/DOCKET NUMBER: 1579-112
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: (703) 816-4000
13 TELEFAX: (703) 816-4100
14 INFORMATION FOR SEQ. ID NO: 6:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 486 amino acids
17 TYPE: amino acid
18 STRANDEDNESS:
19 TOPOLOGY: linear
20 MOLECULE TYPE: protein
21 US-08-684-594-6
22
23 Query Match 68.5%; Score 2302.5; DB 2; Length 486;
24 Best Local Similarity 91.5%; Pred. No. 2.3e-178;
25 Matches 450; Conservative 9; Mismatches 22; Indels 11; Gaps 3
26
27 QY 40 VGSFALLKCGISOGSLSHVDFSVHKRFLIRVHGOGGQSEPEYEDRLSDRGA 99
28 |||||||
29 Db 1 VGSFALLKCGISOGSLSHVDFSVHKRFLIRVHGOGGQSEPEYEDRLSDRGA 60
30 |||||||
31 QY 100 TLALQVTPQDERIFLCQGR--PR--SQEYRIQLRVYKAPPEPNIQVNPGLGIPVNSKE 154
32 |||||||
33 Db 61 TLALQVTPQDERIFLCQGRKLPGRSTASSASTKLRM-----PNIQVNPGLGIPVNSKE 114
34 |||||||
35 QY 155 PEELATGVGNRPPIQVIMYKNGRPLKEKKRVHIQSSQTVESSGLYTLQSLTKAQVYK 214
36 |||||||
37 Db 115 PEELATGVGNRPPIQVIMYKNGRPLKEKKRVHIQSSQTVESSGLYTLQSLTKAQVYK 174
38 |||||||
39 QY 215 EDKDAQFCYELNRLDPSGNHMKESREVTVPYVYPEKFWLEVEPEVGMLEKGDREYIEICLA 274
40 |||||||
41 Db 175 EDKDAQFCYELNRLDPSGNHMKESREVTVPYVYPEKFWLEVEPEVGMLEKGDREYIEICLA 234
42 |||||||
43 QY 275 DGNPPHFSISKONPSTRAEBEETINDNGVLYLEPAKREHSGRYECQGLDITMISLSE 334
44 |||||||
45 Db 235 DGNPPHFSISKONPSTRAEBEETINDNGVLYLEPAKREHSGRYECQGANMIDTMSLSE 294
46 |||||||
47 QY 335 POELLVWYSDVAVSPAAPERQSGSLTLCPEAGSODLEFQWIRETFQVLENGPVLQD 394
48 |||||||
49 Db 295 POELLVWYSDVAVSPAAPERQSGSLTLCPEAGSODLEFQWIRETFQVLENGPVLQD 354
50 |||||||
51 QY 395 HDLKRKRGAGGYRCVASVPSIPGLNRTQVLAIFGPPMAFERKKVYKKNVYLNLSCEA 454
52 |||||||
53 Db 355 HDLKRKRGAGGYRCVASVPSIPGLNRTQVLAIFGPPMAFERKKVYKKNVYLNLSCEA 414
54 |||||||
55 QY 455 SGHRRPITSMNVNGTASBODDQPVYLSLVNLYVLPLELLEGVCTASNDLGKNTISLFL 514
56 |||||||
57 Db 415 SGHRRPITSMNVNGTASBODDQPVYLSLVNLYVLPLELLEGVCTASNDLGKNTISLFL 474
58 |||||||
59 QY 515 ELVNLTLTPPDS 526
60 |||||||
61 Db 475 ELVNLTLTPPDS 486
62 |||||||
63
64 RESULT 3
65 US-08-432-016-2
66 Sequence 2, Application US/08432016
67 Patent No. 5968768
68
69 GENERAL INFORMATION:
70 APPLICANT: HAYNES, BARTON F.
71 APPLICANT: ARUFFO, ALEJANDRO
72 APPLICANT: PATEL, DHAVALKUMAR
73 APPLICANT: BOWEN, MICHAEL A.

```


Db	173	LQYARKEDVYSQTC-----TAKHVMGPDOYSEPESEPIHPIETKVSLOQVYOSQPIR	2255
Qy	264	EGDRVEIRKCLADGNPP--HESISKONPSTREAEETTNDNGVLYLEPARKHSGHYEC	320
Db	226	EGEDVTLTKCQADGNPPPTSFNFNI-----KKKRTVTDKDDVYTLTGTRADSGTYKC	277
Qy	331	QGLDLPMTLSLSEPOELLVNVYDVAVSPAPAE--ROEGSLITLTCEASSODLEFQWLR	379
Db	278	SLLDNDVNES---TQFYATSEFL-DVSLTPGKVLKVNGENLIVSLDKNASSSEKRYWTK	332
Qy	380	EETGCVLVERGEPVLOLHDIKREAGSGGYRCVAVSPISPLNRTOLYVNAIFGPWMA--FKE	437
Db	333	D--NNKLDKLP--DESKLTYS DAGLYCVD--SIEGIKRSLSPELHVEGIPKITSLTKH	385
Qy	438	RKVVWKKEMVNLNLSCEASGHRPPIITSNVNVKTASBQODQPORVLTSLNVLVTPELLETV	497
Db	386	RSSDQKHKY---LTCCEAGSPKPDVOMSVNCTNDEVSYNNKATYKLTLYVPSKNLT--V	439
Qy	498	ECTASNDGKNT 509	
Db	440	SCLYVTKLGEPT 451	

```

1      RESULT      8
2      US-08-684-594-4
3      : Sequence 4, Application US/08684594
4      : Patent No. 5998172
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: HAYNES, BARTON F.
9      : APPLICANT: ARUEFO, ALEJANDRO
10     : APPLICANT: PATEL, DHAVALKUMAR
11     : APPLICANT: BOWEN, MICHAEL A.
12     : APPLICANT: MAROUARDT, HANS
13     : TITLE OF INVENTION: CD6 LIGAND
14     : NUMBER OF SEQUENCES: 14
15     :
16     : CORRESPONDENCE ADDRESSES:
17     :
18     : ADDRESSEE: NIXON & VANDERHAYE P.C.
19     : STREET: 1100 NORTH GLEBE ROAD
20     : CITY: ARLINGTON
21     : STATE: VIRGINIA
22     : COUNTRY: U.S.A.
23     : ZIP: 22201-4714
24     :
25     : COMPUTER READABLE FORM:
26     :
27     : MEDIUM TYPE: Floppy disk
28     : COMPUTER: IBM PC compatible
29     : OPERATING SYSTEM: PC-DOS/MS-DOS
30     : SOFTWARE: PatentIn Release #1.0, Version #1.30
31     :
32     : CURRENT APPLICATION DATA:
33     :
34     : APPLICATION NUMBER: US/08/684,594
35     : FILING DATE: 18-JUL-1996
36     :
37     : CLASSIFICATION: 435
38     :
39     : PRIOR APPLICATION DATA:
40     :
41     : APPLICATION NUMBER: US 08/432,016
42     : FILING DATE: 01-MAY-1995
43     :
44     : PRIOR APPLICATION DATA:
45     :
46     : APPLICATION NUMBER: US 08/333,350
47     : FILING DATE: 02-NOV-1994
48     :
49     : PRIOR APPLICATION DATA:
50     :
51     : APPLICATION NUMBER: US 08/143,903
52     : FILING DATE: 02-NOV-1993
53     :
54     : ATTORNEY/AGENT INFORMATION:
55     :
56     : NAME: WILSON, MARY J.
57     :
58     : REGISTRATION NUMBER: 32,955
59     : REFERENCE/DOCKET NUMBER: 15/9-112
60     :
61     : TELECOMMUNICATION INFORMATION:
62     :
63     : TELEPHONE: (703) 816-4000
64     : TELEFAX: (703) 816-4100
65     :
66     : INFORMATION FOR SEQ ID NO: 4:
67     :
68     : SEQUENCE CHARACTERISTICS:
69     :
70     : LENGTH: 466 amino acids
71     : TYPE: amino acid
72     :
73     : STRANDEDNESS:
74     :
75     :

```

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-684-594-4

```

Query Match	12.38;	Score 414.5;	DB 2;	Length 466;
Best Local Similarity	27.68;	Pred. NO. 1.9e-25;		
Matches 136;	Conservative 91;	Mismatches 200;	Indels 65;	Gaps 21;

```

0Y 41 GSTALLKC--GLSSOGSLSHVDMEFVSHKREKRLIFVHVGOGQ---SBEGBEORLSL 94
Db 2 GETIIVPCNDGCTKKPDG-LIFTKKWYKXDGSPEDDLIVKQAOQDEATSVATDYGKRSVSI 60
0Y 95 ODRGATLALIQVYTODEKIFLCO--GKRPRSGEYRIOLKRYKAPBEPNIOVNPILGIVPNK 153
Db 61 A-ANSSLLIAGSLADQVFCMVAVSFTNTEEYSEVKVHKHKSAPVTK-----NNAK 112
0Y 154 EPE-----EVATCGVRNGPYIPQVIWYKNGRPLKEEKNRHHIOSQVES-SGLYTLQSI 207
Db 113 ELEENKLTQLOEGCEYVANNPRAADIMKKNQTLVDODGKTIIITSTIKDKITGTSSTR 172
0Y 208 LKAOLVKEDKDAOFYCELYNLPLBGSNHK----KKSREYIYVYFPTPKVWLEVEBPVMK 263
Db 173 LQYARKEDVESQFPC-----PAKHMGPDQVSESEPEFIHPTKEVSLQVWOSQPIR 225
0Y 264 EGDREYICGLDGNPNP---HFSSIKQNPSPREAEETNDNGVLVEPARKEHSGRYEC 320
Db 226 EGEDYTLICQADGNPNPISFNFI-----KKKTYLVDKDQVYTLGVTBADSGLKYC 277
0Y 321 OGLDLDLTMISLSEFOELLVNVSDRVYSPAPE-ROBGSLSLITCAESSQDLEFQWL 379
Db 278 SLDDNDVNES---IQFYTVSFL-DVSLTPPGKVLKLVNGEMLIYSLDKNASSSEAKYWTK 322
0Y 380 EFTGQVLRGRPYLOLHDLCREAGGGRVAVSPSLPGNLRIQVLNVALIFGPPMA--FKE 437
Db 333 D--NNKDLKLP--DFSKLTYSDAGLYYCDV---SIEGKNSLSSELYVEGIRKTSLSLKH 385
0Y 438 RRYWVKENNVMLNLSCEASGHPRIITLSMNVNGTASBODODPORVLSTLNVLTVPLELTGV 497
Db 386 RSSDQKHNY---LTCEAGSPRPDVQWGSVNGTNDVEVSNNGNKATYKTLVPSKMLT---V 439
0Y 498 ECTASNDLGAKT 509
Db 440 SCLVYNNKLGEDT 451

```

Query Match	7.5%; Score 252; DB 4; Length 1381;
-------------	-------------------------------------

CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00031
FILING DATE: 19930112
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 715-1000
TELEFAX: (312) 715-1234
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 736 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-00031-15

Query Match 7.1%; Score 237.5; DB 5; Length 736;
Best Local Similarity 20.7%; Pred. No. 8.3e-11;
Matches 146; Conservative 90; Mismatches 251; Indels 217; Gaps 30;

QY 33 PELVEY-----EYGSTALCKGSGSOGNSHVDWFSVHKERTLIFRVGCGQSEP 85
DB 112 PERVELAPLPSPQPVGKNLTLGC---QYEGGAPRRAN-----LTVLLRGKLMKSEF 161
QY 86 GEYEOURLSDRGATLALQTVPDERIFLCOGKR-----PRSGEYRIQLRYKAKE 137
DB 162 LBDADRKSLETKSLEVTFTVEDIGKYLVCRAKLHIDEMSVPTVROAVKELQYISPK 221
QY 138 ENIQVNPPIGIVNSKEPEVATCGRNCPYIPQYIWK---NGRPLMEKKRVIQSSQ 194
DB 222 NTIVISVP--STKLQEGGSVTMTCSSEGLPAPELIFMSKKLDG-----NLOHLSGNA 271
QY 195 T-----VESSGLYTLQSLKAQLVKEDKDAQFYCELYRLPSGNHMKESREVTVPFY 247
DB 272 TLLTLLAMRSDSGIYCEGV---NLI-----GKNRKE-----VELIV 305
QY 248 PREKWLVEPEVGM--KEGDRVELRCLADGNPPPEFSTSKQNPSTREAEETINDNGL 305
DB 306 OEKPFVEIISPGRIMAOIGDSVMLTCSVMGCESEFSFWRTQIDSPLSGKVSSEGTSTL 365
QY 306 VLEPARKEHSRGEC-----OGLDDTMIISLSEPOELVNVYSDVSPAPAPERO--- 356
DB 366 TLPSPVSENEHSTLYCTVTCGKKLEKGIQVELYSFPRD-----PELMSG 410
QY 357 ---EGSSLTLTCEAES---SODLEFQWLREETGOVLERGPIQLDLK----- 398
DB 411 GLVNGSSVTVSCVSPVYPLDRLEIETLLKGET--ILENIEPLEDDIMKLENSKLEMPFI 468
QY 399 ---REGAGGYRCYASV-----PSIPGLNRIQLVNAVIFGPPNMAFERKRYWKENVY 447
DB 469 PTIEDGKALVCOAKLHIDMEFEPRQROSTQTLVYNAV---P---RDTTVLVSPSSI 520
QY 448 L-----NLSCASGHPPTISMNVNGTASEODOPQRYLSTLVNLTPELLETGVE--- 498
DB 521 LERGSVVNMTCTLSQGFPAKILMS-----RQLPNEGLOPLSENATLTLISTKEDSG 572
QY 499 ---CTASNDLGKNTSILFLEL-----VNLTLTLPDS---NTTGLSTSTASPH--- 541
DB 573 VYLCEGINAGRSKRKEVELIIQVTPKDIKLAPPSSEVSKEDPVILISCTCGNVETWILL 632

QY 542 --RANSTYTERK-----LPEPESRGVY----- 561
DB 633 KKAETGDTVLKSIDGAYTIRKAQLKADAGVYECESKNKYGSOLRSLTLDVGRNNKDYF 692
QY 562 ----YAVIYCIYLVAVLGNVLYFLK---KGLPFRSGKOEI 598
DB 693 SPELLVLYFASSLIIPAGIMITYFARKANKGYSVLVEAKSKV 736

RESULT 12
US-09-009-490A-91
Sequence 91, Application US/09009490A
Patent No. 6300491
GENERAL INFORMATION:
APPLICANT: Bennett and Mirabelli
TITLE OF INVENTION: Oligonucleotide Modulation
TITLE OF INVENTION: of Cell Adhesion
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Jane Massey Licata
STREET: 66 East Main Street
CITY: Marlton
STATE: NJ
COUNTRY: USA
ZIP: 08053
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: Windows 95
SOFTWARE: WORDPERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,490A
FILING DATE: January 20, 1998
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 440,740
FILING DATE: May 12, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 063,167
FILING DATE: May 17, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 969,151
FILING DATE: February 10, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 007,997
FILING DATE: January 20, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 939,855
FILING DATE: September 2, 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 567,286
FILING DATE: August 14, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 810-1515
TELEFAX: (609) 810-1454
INFORMATION FOR SEQ ID NO: 91:
SEQUENCE CHARACTERISTICS:
LENGTH: 647
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
US-09-009-490A-91

Query Match 6.9%; Score 233; DB 4; Length 647;
Best Local Similarity 21.9%; Pred. No. 1.6e-10;
Matches 134; Conservative 89; Mismatches 234; Indels 156; Gaps 27;

Query Match 6.9%; Score 232.5; DB 4; Length 1260;
Best Local Similarity 22.4%; Pred. No. 4.8e-10;
Matches 143; Conservative 68; Mismatches 257; Indels 171; Gaps 28;

```
OY 25 PGEAOPAPELVEVEVGSTALLKGLSOGNLSHVDFSVHKEKRTLIIFRVROGQOSE 84
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 PREYVKP-----VEVEGESVVLPCNPPPSAAP-PRITYMN-----SKIFDIKO----- 180
OY 85 PGEYBQRLSDRGATLALTQVTPODERIFLQ-----GKRPSOEYRIQLRVYKAPPEPN 140
    : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 181 ---DERVSMGQNDLYFANVLTSNDHSDYICNAHPGTRILIQEPIDLRV--KPTNSM 234
OY 141 IOVNP-LGIPIVNSKE-----PEEVATCVGRNGYPIPIQVIWYKNGRPLKEEKNRHQIS 192
    | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235 IDRKPRLLFPTNSSRLVALOGOSLIECIA-EGFPPTIKWLHPSDMPD--RVIYON 291
OY 193 SQTVESGLYTLQSLTKAQVLKEDDAQFCELVNRLPDSGNHMKESREVTVPVFPTEKY 252
    | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 292 HNK-----TLOL-----NWGEEDGERTCLAENSLGSARH-----AYTVVEA 330
OY 253 ---MLEVEPVMLKEGDVEIRCLADGNPPPHFSISKQNPSTRAEETINDGVLVLEP 309
    | | : | : | : | : | | | | | | | | | | | | | | |
Db 331 APYMLQKPGSHLYGPGETARLDQVQGRQP-----ETWIRNG-MSMET 374
OY 310 ARKESGRYECOGDL-----DTMISLSEPE--LVN--YVSDRVSPAAPERQ- 356
    | : | | | | | | | | | | | | | | | | | | | |
Db 375 VNKDQRYRIEGLSLISNVQPTDMVTQCEARNQHGILLANAYIVVOLPARILTKDQI 434
OY 357 ---EGSSLTLTCEAESQDLEFQMLRETEGVLE-----RGPVLQHDHKKREAGGY 405
    | | : | | | | | | | | | | | | | | | | | | |
Db 435 YMAVVGSTAYILLCKAFGAPVPSVQMLDEGTTVLQDERFFPYANGTLISIRDLQANDTGRY 494
OY 406 KCVASVPSPISGLNRTQLVNVAIFGPPWMAFERKRWVWKENNVNLSCASGHP--RPTIS 463
    | : | | | | | | : | : | : | : | : | : | : | : |
Db 495 FCOAANDQ---NNVTILANLQVKATQITQGPSAIEKKARVFTQCASFPDLSQASIT 551
OY 464 WNVNCTASEQODQORVLTANLVTPEL---LETGVECTASNDL-----GKN 508
    | : | : | : | : | | | | | | : | | | | |
Db 552 WRGDDRDLOERGDSKYPFIEDGKLVLSQDQNTSCVASTELDEYESRAQLLVGSP 611
OY 509 TSLPLELIVNLTLTPDSNTTGTGSTASPHTR----- 542
    | | | : | | | : | | | | | | | | | | | |
Db 612 GPVPHLEISDRHL---KQSQVHLMSWSPAEDHNSPIEKYDIEFEDKEAPKWFSLGKVP 668
OY 543 ANSISTERKL-----DEPESRGVY 561
    | | | | | | | | | | | | | | | | | |
Db 669 GNOTSTTLKLSPYHYTFRTAINKYGGPESPYSESIV 707
```

Search completed: June 28, 2002, 10:46:59
Job time: 194 sec

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	3327	98.9		646	2	I38049	cell surface glycoprotein
	2	863	25.7		584	2	I50419	s-glycerin precursor
	3	647.5	19.3		628	2	I38000	lutheran blood group
	4	606	18.0		588	2	I37202	B-CAM protein - hu
	5	509.5	15.2		588	2	I40506	adhesion molecule
	6	505.5	15.0		588	2	A45254	surface glycoprotein
	7	502.5	14.9		587	2	JH0464	DM-GRASP precursor
	8	492.5	14.6		583	2	I39428	alcam - human
	9	457	13.6		523	2	I50478	neurofilin - goldfish
	10	281	8.4		5175	2	T20392	hypothetical protein
	11	281	8.4		5186	2	T43390	hemisciental precursor
	12	270	8.0		764	2	A49448	irregular cell adhesion
	13	264.5	7.9		538	2	JC2457	kinase-like protein
	14	259.5	7.7		1051	2	A39712	neural cell adhesion
	15	256	7.6		1088	1	IJXLNLT	perlecan precursor
	16	251.5	7.5		4391	2	A38096	duttil protein
	17	247	7.3		1612	2	T30805	neural cell adhesion
	18	246.5	7.3		725	1	IJMSN5	transmembrane receptor
	19	246.5	7.3		739	2	JS0675	neural cell adhesion
	20	246.5	7.3		1631	2	T14160	heparan sulfate proteoglycan
	21	244.5	7.3		858	1	IJTRNC	neural cell adhesion
	22	244.5	7.3		3707	2	S18252	neural cell adhesion
	23	242.5	7.2		853	1	IJLONC	neural cell adhesion
	24	237	7.0		725	2	JEO099	neural cell adhesion
	25	237	7.0		1092	2	IJN0635	neural cell adhesion
	26	236	7.0		7962	2	I38346	elastic filin - human
	27	235	7.0		1091	1	IJCHNL	neural cell adhesion
	28	234	7.0		761	1	IJHUNG	neural cell adhesion
	29	233.5	6.9		1115	1	IJMSNL	neural cell adhesion

D5 181 LKEKNRVHIQSQTVESSGLYTLQSLKAQLVKEDKDAQFYCELNRYRLPSGNHMKSSRE 240

Db 296 SPSEYTLFR---LQDQEEVLNVNLEGNLTLEGTGSGQGYGCRVEDYDAADVDYLS 351
QY 334 EPQELLVNVSDVRSVPAAPERQEG-----SSLTITCEASSQDLPEQMLRETEGQV 385
Db 352 KTELKRVAYIDPLELS-----EKKVLSLPLNSAVVNCVHGILPTPALMTKDTST--P 402
QY 386 LERGVYQLDHLKREAGGGRVCAVSPSIPGLNRTOLVNAVIFGPPMAFFKRR-----V 440
Db 403 LGGPMLSLSTITFDNSNGTYCEASLPTVPVLSRTQNTFLVQSGPELKTAEIPEKADGS 462
QY 441 WKEMNVLNLSCEASGHRPTISMNVNGTASQEDQDPQR---VLSTLVNLTPELLETGV 497
Db 463 W-REGDEVTLTICSGARGHDPKLSWSQLG-GSPAEPFPGRGWSSSLTKVTSALSROGI 520
QY 498 ECTASNDLGKNTSILFLELVNLTTLTPDSTNTTGLSTASPRTRANSTSTERKLPPEPS 557
Db 521 SCEASNPHGNKRHVHFGAAS-----PQTSQ 546
QY 558 RGVIVAVIYICILVAVLGAVALYFLYKKGLPCRRSGKQETLTPPSRKSSELVEVKSDKL 617
Db 547 AGVAVMAVAVSGLLLLVAVAFYCVRRKGG-PCCRQRRKRGAPP---GEPLSHSGSEQ 602
QY 618 PEEMGLQGG--SSGDKRPRGQGEK 640
Db 603 PEQGTGLMGASGARGSGSGFRGDE 627

RESULT 4
137202
B-CAM protein - human
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
C/Accession: I37202; S47272
R:Campbell, I.G.; Foulkes, M.D.; Senger, G.; Trowsdale, J.; Garin-Chesa, P.; Rettig, W.J.
Cancer Res. 54, 5761-5765, 1994
A:Title: Molecular cloning of the B-CAM cell surface glycoprotein of epithelial cancers:
A:Reference number: I37202; MUID:95042297
A:Accession: I37202
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-588 <RES>
A:Cross-references: EMBL:X80026; NID:9535178; PIDN:CAA56327.1; PID:9535179
A:Genetics:
A:Gene: B-CAM

Query Match 18.0%; Score 606; DB 2; Length 588;
Best local similarity 29.2%; Pred. No. 5.1e-31;
Matches 187; Conservative 96; Mismatches 247; Indels 110; Gaps 25;

QY 2 GLPR-LVCAFLAACCCCPRYAGVPGAEQAPAPLVEVEVSGTALLKCGLSQSGNLSH- 59
Db 12 GAPRLILLAVLLAA-----HPDAQAEVRLSVPLVEVMGKSYILDG---TPGTGHDY 62
QY 60 -VDWF-----SVAKKRTLLFRYRQGGQSGPEBEYDORLSLDRGATLAIYQV 106
Db 63 MLEWFLIDRSGARPLASAEQSGSELOVTMHDTRGRSP-----YQDSQGR--LVLAFA 115
QY 107 TPODERIFLC--QGKRPRSOEYRIQLRYKAPPEPNTIQVPLIGIPVNSKPEEYATQVGR 164
Db 116 QVQGERDYVYCVVRAAGAGTAATARKLAVFAKPEATEVSPKNGTILSWEDSAQELATCNSR 175
QY 165 NGYDIPQVYIWKNGRPK--EKKNRVHIQSSQTV-ESSGLYTLQSLKLAQVLEDKDAQF 221
Db 176 NGNPAPKITWYRNQORLEVEPMNPEGYMTSRVYREASGLISTLSTLYLPCRRDDRRASF 235
QY 222 YCELNLTPLSGNHHK--ESREYTVVFEYTERK--WL--EVEPQMLKEGSRVRELCLADG 276
Db 236 HCAAHYSIPBGRGRDLSPTFHLTLHYPTREHYVFWGSPSTPAGWVEGDTYQLLCRGDG 295
QY 277 NPPHFSISKQNPSTRAEETINDN--GVLVLEPARKHSGRGECGLDIDMTISLSE 334
Db 296 SPSEYTLFR---LQDQEEVLNVNLEGNLTLEGTGSGQGYGCRVEDYDA----- 344

QY 335 PQELLVNVSDVRSVPAAP-ERQEG-----SSLTITCEASSQDLPEQMLRETEGQV 385
Db 345 ADDVQSLKTLVDRAVAYIDPLELSEGGKVLSTPLNSAVVNCVHGILPTPALMTKDTST--P 402
QY 386 LERGVYQLDHLKREAGGGRVCAVSPSIPGLNRTOLVNAVIFGPPMAFFKRR-----V 440
Db 403 LGGPMLSLSTITFDNSNGTYCEASLPTVPVLSRTQNTFLVQSGPELKTAEIPEKADGS 462
QY 441 WKEMNVLNLSCEASGHRPTISMNVNGTASQEDQDPQR---VLSTLVNLTPELLETGV 497
Db 463 W-REGDEVTLTICSGARGHDPKLSWSQLG-GSPAEPFPGRGWSSSLTKVTSALSROGI 520
QY 498 ECTASNDLGKNTSILFLELVNLTTLTPDSTNTTGLSTASPRTRANSTSTERKLPPEPS 557
Db 521 SCEASNPHGNKRHVHFGAAS-----PQTSQ 546
QY 558 RGVIVAVIYICILVAVLGAVALYFLYKKGLPCRRSGKQEG 597
Db 547 AGVAVMAVAVSGLLLLVAVAFYCVRRKGG-PCCRQRRK 585

RESULT 5
JH0506
adhesion molecule SCL precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000
C/Accession: JH0506; PS0270
R:Tanaka, H.; Matsui, T.; Agata, A.; Tomura, M.; Kubota, I.; McFarland, K.C.; Kohr, B.
Neuron 7, 535-545, 1991
A:Title: Molecular cloning and expression of a novel adhesion molecule, SCL.
A:Reference number: JH0506; MUID:92030150
A:Accession: JH0506
A:Molecule type: mRNA
A:Residues: 1-588 <TRAN>
A:Cross-references: GB:S63276; NID:9238000; PIDN:AAB20170.1; PID:9238001
A:Experimental source: embryo
A:Accession: PS0270
A:Molecule type: protein
A:Residues: 34-48 <TRAN>
C:Comment: this protein is uniquely and transiently expressed on spinal cord motoneur
C:Keywords: glycoprotein; transmembrane protein
F:133/Domain: signal sequence #status predicted <SIG>
F:34-588/Product: adhesion molecule SCL #status predicted <ADH>
F:500-523/Domain: transmembrane #status predicted <TRA>
F:101,173,199,271,312,366,462,465,504/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 15.2%; Score 509.5; DB 2; Length 588;
Best local similarity 26.4%; Pred. No. 7.3e-25;
Matches 166; Conservative 117; Mismatches 254; Indels 91; Gaps 30;

QY 6 LVCAFLAACCCCPRYAGVPGAEQAPAPLVEVEVSGTALLKCGLSQSGNLSHYDW-FS 64
Db 18 LILALCMPPALGL-----YTVNAVYGDITMPCRIELVDPD-G-LMFGRMYE 65
QY 65 VHKERTLL-FR--VNOGGQSGPEBEYDORLSLDRGATLAIYQVTPDERIFLCQ--GK 119
Db 66 MPNGSVYFAFSSSTIKKNVQYDVPDYKRLSLSE-NYTLSTKNARISDERKRFVCLVTE 124
QY 120 RPRSOEYRIQLRYKAPPEPNT--QVNPGLIPVNSKPEEYATQVGRNGYPIQVYIYKN 177
Db 125 DVVSEEPY-VKVFQKPSQPELHQADFL---ETEKLMKGEQCVVRSYBGNATWYKN 179
QY 178 GRPLK--EKKNRVHIQSSQTVSSGLYTLQSLKLAQVLEDKDAQFYELNLTPLSGNHM 235
Db 180 GRVLQVEVEVAVINLAKKEN-RSTGLFTMTSLSIQYMPKREANAKFTCTIVYHSGGQKT 238
QY 236 KSRREYTVVFEYTERKVMLEW--EPVGMKKEGSRVRELCLADGNPPH--FSISKQNPST 291
Db 239 IQSEPVYFVNHVPTERTKVTIRVLSQSSITKEGDNVTLKSGNGNPPQGFLEYI----- 291
QY 292 REAEETINDNGVLVLEPARKHSGRGYECGLDIDMTISLSEPQELLVNVSDVRSVPA 351

Db 292 -PGEFEGIRSDTYVMTVRRNNGEYKCSLIDKSM-----DATITIVHL-DLQLPFS 344
Qy 352 AP-ERQEGSSLTITCEASSODLEFQWLREBTGOVLERGVLQHLDKREAGGRCVAS 410
Db 345 GEVTKQIGEALPVSCITSSSRNATVFWIKDNTNR--MKTSP--SFSSLOYODAGNYICETT 400
Qy 411 VPSITGLNRTQLVNVAIFGPPMAFKEKKVWVKNMVLNLSCEASGHPRTISMVNGTA 470
Db 401 LQEVGGLKKRRTKLIVGKRP--QIKMTKKTNTNKNKSKTIVCHVEGFEPKPAQMTVIGSG 458
Qy 471 S--EQDODPQRYLSTLN--VLVTPPELLETGVECTASNDLGKNTLSLELVNLTLPDPS 526
Db 459 SLINKTEETKYVNGKFSSKIIIAPEENVY-LTCLIAENEL-----E 497
Qy 527 NTTGLSTSTAS-----PHTRANSTSTERKLPEPESRGVIVAVYICILVALVAGVLY 580
Db 498 RTVTSILNVSATISPEYDEPDR-NDNSEK----VNDQAKLIVIGVGLLVALVAGVY 552
Qy 581 FLY-KKGKLPGRSGKQETILPPSRKSE 607
Db 553 WLTVKSKTSKXKHVDKLGNIENKKE 580

RESULT 6
A45254
surface glycoprotein BEN precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C:Accession: A45254; S19202
R:Pourquie, O.; Corbel, C.; Le Caer, J.P.; Rossier, J.; Le Douarin, N.M.
Proc. Natl. Acad. Sci. U.S.A. 89, 5261-5265, 1992
A:Title: BEN, a surface glycoprotein of the immunoglobulin superfamily, is expressed in
A:Reference number: A45254; MUID:92302224
A:Accession: A45254
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-588 <POD>
A:Cross-references: EMBL:X64301; NID:963087; PIDN:CMA45579.1; PID:963088
C:Keywords: glycoprotein

Query Match 15.0%; Score 505.5; DB 2; Length 588;
Best Local Similarity 26.4%; Pred. No. 1.3e-24;
Matches 166; Conservative 116; Mismatches 255; Indels 91; Gaps 30;
Qy 6 LYCAFLLAACCCPRVAGPEAEQAPAPLVEVGSFALLKCGLSQSGNLSHVDW-FS 64
Db 18 LLC--LLALALCMPPALGL-----YTVNAVYGDITIMPCRLVDPDG-LMFGKMYE 65
Qy 65 VHKEKRTLI-FR--VRQGGQSGEPGEYEQRLSLQDRGATLALTQVTPDERIFLCO--GK 119
Db 66 MNGSPVFIARFSSTKKNQYDDVDYDKRSLSE--NTLSIKNARISDEKRFVCMLYTE 124
Qy 120 RPRSQEYRIQLRVYKAREEPNI--QVNPGLIPVNSKEPEVATCGRNGYPIPOVIWYKN 177
Db 125 DVSEEPYV-KVYFKQSPQPEILHQADF-----ETEKLMGCECVARSYPGENTWYKN 179
Qy 178 GRPLK--EKNRVHIQSSQTVESGGLTYLQSLKQALVQEDAOFCYELNRLPSGNHM 235
Db 180 GRVLQVEEYVAVINLKKVYN-RSTGLFTMTSSLOYMPTEKEDNAKFCIYVTHGSGOKT 238
Qy 236 KESRETVVPEYPTKEKWLVEV-EPVGMLEKGGDRVEIRCLADGNPPH--FSISKONPST 291
Db 239 IQSEPFVDFVHYPTKEVTRIVLSQSTIKEGNNTLKCGNGNPNPQPELFYI----- 291
Qy 292 REAEETITNDNGVILVLEPARKHSGRYECQGLDITMISLSEPOELLVNYSDVRVSPA 351
Db 292 -PGETEGIRSDTYVMTVRRNNGEYKCSLIDKSM-----DTTITIVHYL-DLQLPFS 344
Qy 352 AP-ERQEGSSLTITCEASSODLEFQWLREBTGOVLERGVLQHLDKREAGGRCVAS 410
Db 345 GEVTKQIGEALPVSCITSSSRNATVFWIKDNTNR--MKTSP--SFSSLOYODAGNYICETT 400

Qy 411 VPSITGLNRTQLVNVAIFGPPMAFKEKKVWVKNMVLNLSCEASGHPRTISMVNGTA 470
Db 401 HKEVEGLKKRRTKLIVGKRP--QIKMTKKTNTNKNKSKTIVCHVEGFEPKPAQMTVIGSG 458
Qy 471 S--EQDODPQRYLSTLN--VLVTPPELLETGVECTASNDLGKNTLSLELVNLTLPDPS 526
Db 459 SLINKTEETKYVNGKFSSKIIIAPEENVY-LTCLIAENEL-----E 497
Qy 527 NTTGLSTSTAS-----PHTRANSTSTERKLPEPESRGVIVAVYICILVALVAGVLY 580
Db 498 RTVTSILNVSATISPEYDEPDR-NDNSEK----VNDQAKLIVIGVGLLVALVAGVY 552
Qy 581 FLY-KKGKLPGRSGKQETILPPSRKSE 607
Db 553 WLTVKSKTSKXKHVDKLGNIENKKE 580

RESULT 7
JH0464
DM-GRASP precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 02-Sep-2000
C:Accession: JH0464
R:Burns, F.R.; von Kannen, S.; Guy, L.; Raper, J.A.; Kamholz, J.; Chang, S.
Neuron 7, 209-220, 1991
A:Title: DM-GRASP, a novel immunoglobulin superfamily axonal surface protein that sup
A:Reference number: JH0464; MUID:91337449
A:Accession: JH0464
A:Molecule type: mRNA
A:Residues: 1-587 <BUR>
A:Experimental source: brain
C:Comment: This is a cell surface glycoprotein.
C:Keywords: glycoprotein
F:1-32/Domain: signal sequence #status predicted <SIS>
F:33-367/Product: DM-GRASP #status predicted <DM>
F:67,198,270,311,365,461,484,503/Binding site: carbohydrate (asn) (covalent) #status

Query Match 14.9%; Score 502.5; DB 2; Length 587;
Best Local Similarity 26.3%; Pred. No. 2e-24;
Matches 165; Conservative 118; Mismatches 254; Indels 91; Gaps 30;
Qy 6 LYCAFLLAACCCPRVAGPEAEQAPAPLVEVGSFALLKCGLSQSGNLSHVDW-FS 64
Db 17 LLC--LLALALCMPPALGL-----YTVNAVYGDITIMPCRLVDPDG-LMFGKMYE 64
Qy 65 VHKEKRTLI-FR--VRQGGQSGEPGEYEQRLSLQDRGATLALTQVTPDERIFLCO--GK 119
Db 65 MNGSPVFIARFSSTKKNQYDDVDYDKRSLSE--NTLSIKNARIRHKKRFVCMLYTE 123
Qy 120 RPRSQEYRIQLRVYKAREEPNI--QVNPGLIPVNSKEPEVATCGRNGYPIPOVIWYKN 177
Db 124 DVSEEPYV-KVYFKQSPQPEILHQADF-----ETEKLMGCECVARSYPGENTWYKN 178
Qy 178 GRPLK--EKNRVHIQSSQTVESGGLTYLQSLKQALVQEDKQAOFCYELNRLPSGNHM 235
Db 179 GRVLQVEEYVAVINLKKVYN-RSTGLFTMTSSLOYMPTEKEDNAKFCIYVTHGSGOKT 237
Qy 236 KESRETVVPEYPTKEKWLVEV-EPVGMLEKGGDRVEIRCLADGNPPH--FSISKONPST 291
Db 238 IQSEPFVDFVHYPTKEVTRIVLSQSTIKEGNNTLKCGNGNPNPQPELFYI----- 290
Qy 292 REAEETITNDNGVILVLEPARKHSGRYECQGLDITMISLSEPOELLVNYSDVRVSPA 351
Db 291 -PGETEGIRSDTYVMTVRRNNGEYKCSLIDKSM-----DATITIVHYL-DLQLPFS 343
Qy 352 AP-ERQEGSSLTITCEASSODLEFQWLREBTGOVLERGVLQHLDKREAGGRCVAS 410
Db 344 GEVTKQIGEALPVSCITSSSRNATVFWIKDNTNR--MKTSP--SFSSLOYODAGNYICETT 399
Qy 411 VPSITGLNRTQLVNVAIFGPPMAFKEKKVWVKNMVLNLSCEASGHPRTISMVNGTA 470

C:Accession: T20992; T24733
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20992
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5175 <W12>
A:Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24733
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5175 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a
C:Genetics:
A:Gene: CESP:F15G9.4a
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
A:2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/7;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5077/1

Query Match 8.4%; Score 281; DB 2; Length 5175;
Best Local Similarity 22.1%; Pred. No. 5.6e-09;
Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAEOPAPELVEVVGSTAL-----LKCGLSOSQGNLSHVDFSVYKKEKRTLIFRVG- 79
Db 1357 DVQEPPIILPSTQNTNTAVGVDELKCYVEASPP--ASVYWF-----RRGI 1401
QY 80 -OGOSEPEYEORLSLODRGATLALQVTPDERIFLCQGRPRSQ--EYRQLRVYKAP 137
Db 1402 AIGDTKRG-----YVESDGLVIOASVEDATITTCASNPAGKAEANLQVTVIASPD 1455
QY 138 --EPNIQVNPGLIPVNSKPEEVATCVGRNGYPIPOVIMYKNGRPL-----KEKN 186
Db 1456 IKDPDVVTOE---SIKESHPFSLYCPVFSN--PLQISWLYNDKPLIDDKTSWKTSDKR 1510
QY 187 RVHIQSSQTVESGGLYTLQSIILKAQLYKEDDAQFCYCELNRLPFGNHNKESR--EVTVPV 245
Db 1511 KLHVFKAK--ITDSGVYKCAV-----RNAAGGSKSQVYEVIYPL 1548
QY 246 FYPTKQVWLEVEPVGMLEKGRDVEIRCLADGNPPPHFSI---SKQNPSTREAEETTND 301
Db 1549 NLDSKYYKKV---FAKGEVETLGCVPVSGFPVQIMWVVDGTVVEPGKKYKATLSND 1604
QY 302 NGVLYLEPARKHSGRYECOG-----LDLDTMISLSEPOELLVNYSDVRVSPAAPER 355
Db 1605 GLTLHFDSVSVKQEGNYHCVAQSKGNLIDIVELSVLAVP---IVGEDNLEVF----- 1655
QY 356 QEGSSLTITC--EAESSODLEFOWL-----REETGOVLERGVLQDLHLKREAGGGR 406
Db 1656 -LGDISTISCDLQTESDDKTTFTVWSINGESDRPDNVOIIPSDGRLITITAKPPNNKKY 1714
QY 407 CVASVPSTIGLNRQTLQVNV--AIFGPPMAFKERKVVKNMVLNLSCEASGHPRTIS 463
Db 1715 CRVY--NSAGKAKERTLTDVLEPVPFVEP--VFEANQKLIIGNNPII--LQCVTGPKRTVI 1770
QY 464 WNVGTSSEODQDQRYLSTLNLVLPPELLTG---VECTASNDLGKNTSIFLELVNL 519
Db 1771 WKIDGNVDKSMLEDESLSLIRI---EKL--TKSKAQISCTAEKAKGTASRDFFIQIAA 1825
QY 520 TLTFFDSNTTT 530
Db 1826 PTFKNEGDQET 1836

RESULT 11

T43290
hemiscentin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
A:Accession: T43290; T20993; T24734
R:Voegel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A:Description: Hemiscentin is required for hemidesmosome mediated cell adhesion and ge
A:Reference number: Z22396
A:Accession: T43290
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <VOG>
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20993
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone T0989
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184
A:2512/2; 2593/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 303
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

Query Match 8.4%; Score 281; DB 2; Length 5198;
Best Local Similarity 22.1%; Pred. No. 5.6e-09;
Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAEOPAPELVEVVGSTAL-----LKCGLSOSQGNLSHVDFSVYKKEKRTLIFRVG- 79
Db 1357 DVQEPPIILPSTQNTNTAVGVDELKCYVEASPP--ASVYWF-----RRGI 1401
QY 80 -OGOSEPEYEORLSLODRGATLALQVTPDERIFLCQGRPRSQ--EYRQLRVYKAP 137
Db 1402 AIGDTKRG-----YVESDGLVIOASVEDATITTCASNPAGKAEANLQVTVIASPD 1455
QY 138 --EPNIQVNPGLIPVNSKPEEVATCVGRNGYPIPOVIMYKNGRPL-----KEKN 186
Db 1456 IKDPDVVTOE---SIKESHPFSLYCPVFSN--PLQISWLYNDKPLIDDKTSWKTSDKR 1510
QY 187 RVHIQSSQTVESGGLYTLQSIILKAQLYKEDDAQFCYCELNRLPFGNHNKESR--EVTVPV 245
Db 1511 KLHVFKAK--ITDSGVYKCAV-----RNAAGGSKSQVYEVIYPL 1548
QY 246 FYPTKQVWLEVEPVGMLEKGRDVEIRCLADGNPPPHFSI---SKQNPSTREAEETTND 301
Db 1549 NLDSKYYKKV---FAKGEVETLGCVPVSGFPVQIMWVVDGTVVEPGKKYKATLSND 1604
QY 302 NGVLYLEPARKHSGRYECOG-----LDLDTMISLSEPOELLVNYSDVRVSPAAPER 355
Db 1605 GLTLHFDSVSVKQEGNYHCVAQSKGNLIDIVELSVLAVP---IVGEDNLEVF----- 1655
QY 356 QEGSSLTITC--EAESSODLEFOWL-----REETGOVLERGVPYQLHLKREAGGGR 406
Db 1656 -LGDISTISCDLQTESDDKTTFTVWSINGESDRPDNVOIIPSDGRLITITAKPPNNKKY 1714
QY 407 CVASVPSTIGLNRQTLQVNV--AIFGPPMAFKERKVVKNMVLNLSCEASGHPRTIS 463

Db 1715 CRVT-NSAGKAEKRTLLTLLVLEPPVEFE--VEEANOQKLIIGNPIT-LOCCQVTGNPKPTVI 1770
QY 464 MNVNSTASEODODPQRLVSTLNVLTPPELLTEG----VECTASNDLGNSTILFLEVLNL 519
Db 1771 WKIDSDNDVKSMLPDESLSLIRI-----EKL-TGKSAQISCTAENKAGTASHDFTIOMIAA 1825
QY 520 TTLPPDSNTTT 530
Db 1826 PTFKNEGDEET 1836

RESULT 12

A49448
Irregular chiasm C-roughnest precursor - fruit fly (Drosophila melanogaster)
N:Alternate names: Irrec-roughnest protein
C:Species: Drosophila melanogaster
C>Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 07-May-1999
C:Accession: A49448; S34129
R:Ramos, R.G.P.; Igloi, G.L.; Lichte, B.; Baumann, U.; Mäler, D.; Schneider, T.; Brandst
Genes Dev. 7, 2533-2547, 1993
A:Title: The irregular chiasm C-roughnest locus of Drosophila, which affects axonal proje
A:Reference number: A49448; MUID:94102535
A:Accession: A49448
A>Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-764 <RAM>
A:Cross-references: GB:L11040; NID:g304790; PID:g304791; EMBL:Z21641; NID:g312985; PID:g
C:Genetics:
A:Gene: FlyBase:rst
A:Cross-references: FlyBase:FBgn0003285
C:Keywords: transmembrane protein

Query Match 8.0%; Score 270; DB 2; Length 764;
Best Local Similarity 20.7%; Pred. No. 2e-09;
Matches 137; Conservative 89; Mismatches 277; Indels 158; Gaps 25;

QY 40 VGSTALLKCGLSOSQGNLSHV-DMFSVHKERKRTLIFFRVROGQSGEPGEYFORLSIDRG 98
Db 41 VGARVTLPCRVYINKGQTLQWTKDQGLGTSRDLGFEHYANVGSDEGDY----- 90
QY 99 ATLALTOVTPDDEKIFLCQK-----RPRSOEYRIQLRYKAPRPNIQVNLGIPVNSK 153
Db 91 -SLDIYPMALDDARQCQVSPGEGPAPRISTFAGLTIVLPAPAPKITQGDVIYATADR 149
QY 154 EPEEVATCGVNGYPIPOVITWYKN-GR-----PLKEKNRHHIQQSVYESSGL 201
Db 150 KVE--IECVSVGKRAAETITWIDGLGNVLNDIEYIVIPLPDQR----- 192
QY 202 YTLQSLKLAQLVKEDKDAQFYCELNRYLRPSGNHMKESREYVVPFYPTKRWLEVEPVGM 261
Db 193 FTASVLTLPFKKEHNNTFSCQAQ--NTADRTYRSKRIVEVKYAPK--YAVNNMGS 246
QY 262 LKEG-----DRVEINCLADGNPP--HFSISKQNPSTREAE 295
Db 247 LPGAAGSVGAGAGSVHMSYTSRIVESHQVRLERADANDSDVRYWFIINDEPILIGQK 306
QY 296 EETNDNGVLVLEPARKHSRREGQGLDPTMISLSEPELLVNVYSDVRSPAPAPR 355
Db 307 TE-----MVRNVTYRKHDAIVKC--EYQNSVGSSESDSTLDTISTAPSRQRPQSEMA 357
QY 356 QEGSSLITLCEASSODLEFQMLRETEGOVLERGPVLQDLHLKREAGGVCYVSPISIP 415
Db 358 DVGSVSLTLCVDSNPQEIYWIQHPSDRYVGTSTNLTF--SVSNETAGRYTCRANVGYA 416
QY 416 GLNRQLVNVALIFGPPMAFEKRYVWKENVYLMSCASGHPPT--ISNVNG--TASE 472
Db 417 EISADAVYLLK--GSPAIG--SQRTQYGLVGTARIIECFASSVPARAHVSWTFNGQETISE 473
QY 473 QDOD-----PQVLTSTLNVLTPPELLTEGECTASNDLGNSTILFLEVLNLTLTP 524
Db 474 SGHDYSILYDAVPGGVKSTLIIRDSQAIHYGKYNCYVNDIGNDVAETLOAKKSVSL-- 531

QY 525 DSNVTGISTSTASPHTRANSTSTERKLPESREGVYIVAVIYCIJVLAVLGYFLTK 584
Db 532 -----LMTIVG-----ISVAFLLVLTIL-VVY----- 555
QY 585 KGKLPGRSGKQETLTPPSKSELVYEVKSKLPEEMGLQSSGDKRAPGDQGERYIDL 644
Db 556 ---IKCKKRTK---LPP-----ADVISEHQITRNKGVSCKLEPGRGRTSNYSDL 597
QY 645 R 645
Db 598 K 598

RESULT 13

JC2457
Vascular cell adhesion protein - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 15-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 08-Oct-1999
C:Accession: JC2457
R:Tsang, Y.T.M.; Haskard, D.O.; Robinson, M.K.
Biochem. Biophys. Res. Commun. 201, 805-812, 1994
A:Title: Cloning and expression kinetics of porcine vascular cell adhesion molecule.
A:Reference number: JC2457; MUID:94271236
A:Accession: JC2457
A:Molecule type: mRNA
A:Residues: 1-538 <TSA>
A:Cross-references: EMBL:008351; NID:g474382; PID:AAA21542.1; PID:g474383
C:Keywords: glycoprotein; transmembrane protein
F:497-517/Domain: transmembrane status predicted <TM>
F:75,157,271,330,360/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.9%; Score 264.5; DB 2; Length 538;
Best Local Similarity 21.6%; Pred. No. 2.8e-09;
Matches 126; Conservative 103; Mismatches 234; Indels 119; Gaps 27;

QY 31 PABEIVEVEGSTALLKCGLSOSQGNLSHVDMFSVHKERKRTLIFFRVROGQSGEPGEYEQ 90
Db 31 PEKMT-AQIGDASLSLTCAPDCESSLS-FSW-----RTQIDSPNG----- 70
QY 91 RLSLDQSGATLALTOVTPDDEKIFLC-----QGRPRSOEYRIQLRYKAPRPNIQV 143
Db 71 KVTNGTRSTLVNPNVSPFENEHSYLTCTVSCGNLKGER-----GIQVETISPPKPEIHW 124
QY 144 NPL---GIPIVNSKEPEVATCVGRNGYPIR--QYIWKNGRGLKKEK--NRHHIQQSV 196
Db 125 SSLPEVGRPYTR-----CLVPDVPVPEKLEIELLNKNSHMSVQNFLELIDISKETK 177
QY 197 ESSGLYTLQSLKLAQLVKEDKDAQFYCELNRYLRPSGNHMKESREYVVPFYPTKRWLEV 256
Db 178 SLEFTT-----PTFEDIGAIYCAQATLLIDQSPSKYITPEKQYIISKDVP-ISV 228
QY 257 EYVGMLEKGDVREIRCLADGNPPRPHFSISKQNPSTREAEETNDNGVLVLEPARKHS 316
Db 229 NPSTLSQEGSNMNTCTSEGLPAPQISWGR--LDNGQQLLSGATLTLTLLAMEDESG 285
QY 317 RYECQGLDPTMISLSEPELLVNVY---SDVRSPAPRQEGSSSLITLCEASSODL 373
Db 286 IYVCEGVN--PVGNRKEVELTVQAPKDTTISVNPSS-TLEEGSVMTQSSSGFPAP 341
QY 374 EFQW---LRETEGOVLERGPVLQDLHLKREAGGVCYVSPISGLNTOVLVNAIFGP 430
Db 342 KILMSKRLRGNEIPESENTTLITLSTKNEDSGIYVCEBI--NQGINKKE-VELIQA 398
QY 431 P-----WMAFERKRVWYKENVYLMSCASGHPPTISMVNVGTASBQDODPQVISTLV 486
Db 399 PKDLQTLAPPSVSKGQDVIIISCTC--GNVPPTLI-----ILKKAETGPTVLKSTDG 450
QY 487 LVT---PELLEGV--ECTASNDLGNSTILFLEVLNLTLTPDSNTTGLSTSTASPHTR 542
Db 451 AYTIHRRLADAGVYECESKNEIG-----LQLRSTITLDVKGRESN----- 490

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:46:35 ; Search time 14.84 Seconds

(without alignments)
1585.501 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363
Sequence: 1 MGLPRLVCAFLAACCPCPR.....SSGDKRAPGDGGEKTYDLRH 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3327	98.9	646	1	MU18_HUMAN
2	647.5	19.3	628	1	LU_HUMAN
3	509.5	15.2	588	1	C166_CHICK
4	492.5	14.6	583	1	C166_HUMAN
5	488.5	13.5	583	1	C166_MOUSE
6	465.5	13.8	553	1	C166_CARAU
7	438	13.0	564	1	C166_BRARE
8	270	8.0	764	1	ICCR_DROME
9	259.5	7.7	1051	1	PKK7_CHICK
10	256	7.6	1088	1	NCA1_CHICK
11	246.5	7.3	725	1	NCA2_XENLA
12	246.5	7.3	739	1	NCA1_RAT
13	244.5	7.3	858	1	NCA1_MOUSE
14	244.5	7.3	3707	1	PGBM_MOUSE
15	242.5	7.2	853	1	NCA1_BOVIN
16	237	7.0	1092	1	NCA2_XENLA
17	235.5	7.0	1093	1	PGBM_HUMAN
18	235	7.0	4991	1	NCA1_HUMAN
19	234	7.0	761	1	NCA2_HUMAN
20	234	6.9	848	1	NCA1_HUMAN
21	233.5	6.9	1115	1	NCA1_MOUSE
22	232.5	6.9	1260	1	NCA1_MOUSE
23	230	6.8	739	1	VCA1_MOUSE
24	230	6.8	837	1	NCA2_MOUSE
25	228	6.8	811	1	FS22_DROME
26	228	6.8	873	1	FS21_DROME
27	226.5	6.7	1259	1	CAML_RAT
28	226	6.7	738	1	PECL_HUMAN
29	225.5	6.7	1257	1	CAML_HUMAN
30	225	6.7	702	1	CEAS_HUMAN
31	224.5	6.7	1070	1	PKK7_HUMAN
32	218.5	6.5	404	1	RAGE_HUMAN
33	217.5	6.5	1709	1	SN_HUMAN

34	215	6.4	1284	1	NRCA_CHICK	P35331	gallus gall
35	214	6.4	847	1	CD22_HUMAN	P20273	homo sapien
36	210	6.2	1333	1	VGR1_MOUSE	P35969	mus musculus
37	209	6.2	862	1	CD22_MOUSE	P35329	mus musculus
38	208	6.2	739	1	VCA1_MOUSE	P29533	mus musculus
39	207	6.2	526	1	CEAL_HUMAN	P13688	homo sapien
40	205.5	6.1	402	1	RAGE_HUMAN	P35969	rattus norv
41	205.5	6.1	1356	1	VGR2_HUMAN	P35968	homo sapien
42	204.5	6.1	416	1	RAGE_BOVIN	Q28173	bos taurus
43	204.5	6.1	2012	1	NSCA_HUMAN	O60469	homo sapien
44	203.5	6.1	1266	1	NSCA_CHICK	O03696	gallus gall
45	203	6.0	1447	1	DCC_MOUSE	P70211	mus musculus

ALIGNMENTS

RESULT	ID	MU18_HUMAN	STANDARD	PRT	646 AA.
AC	PA3121				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Cell surface glycoprotein MUC18 precursor (Melanoma-associated antigen MUC18) (Melanoma-associated antigen A32) (S-endo 1 endothelial-associated antigen) (CD146 antigen) (Melanoma adhesion molecule).				
DE	associated antigen (CD146 antigen) (Melanoma adhesion molecule).				
GN	MCAM OR MUC18.				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RC	TISSUE=Melanoma;				
RX	MEDLINE=90093368; PubMed=2602381;				
RA	Lehmann J.M., Rietmueller G., Johnson J.P.;				
RT	"MUC18, a marker of tumor progression in human melanoma, shows				
RT	sequence similarity to the neutral cell adhesion molecules of the				
RT	immunoglobulin superfamily.";				
RT	Proc. Natl. Acad. Sci. U.S.A. 86:9891-9895(1989).				
RL	[2]				
RN	REVISIONS, SEQUENCE FROM N.A.				
RP	TISSUE=Melanoma;				
RC	MEDLINE=93391384; PubMed=8378324;				
RX	Sers C., Kirsch K., Rothbacher U., Rietmueller G., Johnson J.P.;				
RA	"Genomic organization of the melanoma-associated glycoprotein MUC18:				
RT	implications for the evolution of the immunoglobulin domains.";				
RT	Proc. Natl. Acad. Sci. U.S.A. 90:8514-8518(1993).				
RL	[3]				
RN	SEQUENCE OF 24-44; 98-112; 135-153; 240-260; 379-389 AND 460-478.				
RP	MEDLINE=9415196; PubMed=8162602;				
RX	Shih I.-M., Elleder D.E., Speicher D., Johnson J.P., Herlyn M.;				
RA	"Isolation and functional characterization of the A32 melanoma-				
RT	associated antigen.";				
RT	Cancer Res. 54:2514-2520(1994).				
RL	[4]				
RN	SEQUENCE OF 27-40; 98-112 AND 236-260.				
RP	MEDLINE=96136302; PubMed=8573133;				
RX	Bardin N., Frances V., Lesaulle G., Horschowski N., George F.,				
RA	Sampol J.;				
RT	"Identification of the S-Endo 1 endothelial-associated antigen.";				
RT	Biochem. Biophys. Res. Commun. 218:210-216(1996).				
RL	[5]				
RP	FUNCTION.				
RX	MEDLINE=94122526; PubMed=8292890;				
RA	Johnson J.P., Rothbacher U., Sers C.;				
RT	"The progression associated antigen MUC18: a unique member of the				
RT	immunoglobulin supergene family.";				
RL	Melanoma Res. 3:337-340(1993).				
CC	-I- FUNCTION: COULD BE AN ADHESION MOLECULE ACTIVE IN NEURAL CREST				
CC	CELLS DURING EMBRYONIC DEVELOPMENT. ITS EXPRESSION MAY ALLOW				
CC	MELANOMA CELLS TO INTERACT WITH CELLULAR ELEMENTS OF THE VASCULAR				

CC SYSTEM THEREBY ENHANCING HEMATOGENOUS TUMOR SPREAD.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC TISSUE SPECIFICITY: MAY APPEAR AT THE SURFACE OF NEURAL CREST
 CC CELLS DURING THEIR EMBRYONIC MIGRATION. APPEARS TO BE LIMITED TO
 CC VASCULAR SMOOTH MUSCLE IN NORMAL ADULT TISSUES. ASSOCIATED WITH
 CC TUMOR PROGRESSION AND THE DEVELOPMENT OF METASTASIS IN HUMAN
 CC MALIGNANT MELANOMA. EXPRESSED MOST STRONGLY ON METASTATIC LESIONS
 CC AND ADVANCED PRIMARY TUMORS AND IS ONLY RARELY DETECTED IN BENIGN
 CC MELANOCYTIC NEVI AND THIN PRIMARY MELANOMAS WITH A LOW PROBABILITY
 CC OF METASTASIS.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC -1- DATABASE: NAME-PROW; NOTE-CD guide CD146 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd146.htm"
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M29277; AAA20824.1; -
 CC EMBL: M28882; AAA20922.1; -
 CC EMBL: X68264; CAA48332.1; -
 CC EMBL: X68265; CAA48332.1; JOINED.
 CC EMBL: X68266; CAA48332.1; JOINED.
 CC EMBL: X68267; CAA48332.1; JOINED.
 CC EMBL: X68268; CAA48332.1; JOINED.
 CC EMBL: X68270; CAA48332.1; JOINED.
 CC EMBL: X68271; CAA48332.1; JOINED.
 CC MIM: 155735; -
 CC InterPro: IPR003006; Ig_MHC.
 CC InterPro: IPR003598; Ig_C2.
 CC InterPro: IPR003600; Ig_Like.
 CC Pfam: PF00047; Ig; 5.
 CC SMART: SM00410; Ig_Like; 2.
 CC SMART: SM00408; IGC2; 2.
 CC Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;
 CC Repeat; Signal.
 CC SIGNAL 1 23
 CC CHAIN 24 646 CELL SURFACE GLYCOPROTEIN MUC18.
 CC DOMAIN 24 559 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 560 583 POTENTIAL.
 CC DOMAIN 584 646 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 39 122 IG-LIKE V-TYPE DOMAIN 1.
 CC DOMAIN 154 229 IG-LIKE V-TYPE DOMAIN 2.
 CC DOMAIN 265 327 IG-LIKE C2-TYPE DOMAIN 1.
 CC DOMAIN 358 414 IG-LIKE C2-TYPE DOMAIN 2.
 CC DOMAIN 445 506 IG-LIKE C2-TYPE DOMAIN 3.
 CC DISULFID 48 116 PROBABLE.
 CC DISULFID 161 223 PROBABLE.
 CC DISULFID 272 320 PROBABLE.
 CC DISULFID 365 407 PROBABLE.
 CC DISULFID 452 499 PROBABLE.
 CC CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 508 508 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 646 AA; 71793 MW; F064A5DABE0BADE6 CMC64;

Query Match 98.9%; Score 3327; DB 1; Length 646;
 Best Local Similarity 98.9%; Pred. No. 3.4e-212;
 Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGLPLVCAFLIAACCCPRVAGVGEAEQAPAEVVEVSTALLKGLGSLQSGNLSHV 60

DB 1 MGLPLVCAFLIAACCCPRVAGVGEAEQAPAEVVEVSTALLKGLGSLQSGNLSHV 60
 QY 61 DMFSVAKERTLIFRVQGGSEGEVEQRLSLQDRATLALQVYQDRIFLCOGKR 120
 DB 61 DMFSVAKERTLIFRVQGGSEGEVEQRLSLQDRATLALQVYQDRIFLCOGKR 120
 QY 121 PPSQETRIQRYKAPPEEINQVNLGIPVNSKEPEEATCGRNGYIPQVYWKNGRP 180
 DB 121 PPSQETRIQRYKAPPEEINQVNLGIPVNSKEPEEATCGRNGYIPQVYWKNGRP 180
 QY 181 LEEKNRVHIQSSQTVESGLTYLQSLKALVKEDKDAQFCYELNTPSGNHHKESRE 240
 DB 181 LEEKNRVHIQSSQTVESGLTYLQSLKALVKEDKDAQFCYELNTPSGNHHKESRE 240
 QY 241 VYVYVYPTREKWLVEEPYGMKEGDRVIRCLADGNPPHSISKONPSTREAEETTN 300
 DB 241 VYVYVYPTREKWLVEEPYGMKEGDRVIRCLADGNPPHSISKONPSTREAEETTN 300
 QY 301 DNGVLVLEPARKHSGRYECQGLDPTMTLSLSEPOELLVNYSDRVSPAAPERQEGSS 360
 DB 301 DNGVLVLEPARKHSGRYECQGLDPTMTLSLSEPOELLVNYSDRVSPAAPERQEGSS 360
 QY 361 LITTCABESSQDLFEQWLEEFQVLERGPVQLHDKREAGGYRCAVSPSTGLNRT 420
 DB 361 LITTCABESSQDLFEQWLEEFQVLERGPVQLHDKREAGGYRCAVSPSTGLNRT 420
 QY 421 QLYNVAIFGPYMAFKERYWYKENVNLNLSCEASGHPRTISMVNYGTASBODDPQRY 480
 DB 421 QLYNVAIFGPYMAFKERYWYKENVNLNLSCEASGHPRTISMVNYGTASBODDPQRY 480
 QY 481 LSTLVNLVLPPELLTGVECTASNDLCKNTSIFLELVNLTITLTPDSNTTGTSTASGH 540
 DB 481 LSTLVNLVLPPELLTGVECTASNDLCKNTSIFLELVNLTITLTPDSNTTGTSTASGH 540
 QY 541 TRANSTSTERKLPPEPSRGVYAVYVLCITLVAVGAVYFYKKGKLPGRSGKQETL 600
 DB 541 TRANSTSTERKLPPEPSRGVYAVYVLCITLVAVGAVYFYKKGKLPGRSGKQETL 600
 QY 601 PPSRSELYVEVSKDKLPPEEMGLQSSGDKRAPOGGEKYIDLRLH 646
 DB 601 PPSRSELYVEVSKDKLPPEEMGLQSSGDKRAPOGGEKYIDLRLH 646
 RESULT 2
 LU HUMAN STANDARD; PRT; 628 AA.
 ID LU HUMAN
 AC P50895;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Lutheran blood group glycoprotein precursor (B-CAM cell surface
 DE glycoprotein) (Auburger B antigen) (F8/G253 antigen).
 GN LU OR BCM OR MSK19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-67 AND 182-203.
 RC TISSUE=Placenta;
 RX MEDLINE=95296337; PubMed=7777537;
 RA Parsons S.F., Mallinson G., Holmes C.H., Houlihan J.M., Simpson K.L.,
 RA Mawby W.J., Spurr N.K., Warne D., Barclay A.N., Anstee D.J.,
 RA "The Lutheran blood group glycoprotein, another member of the
 RT immunoglobulin superfamily, is widely expressed in human tissues and
 RT is developmentally regulated in human liver".
 RL Proc. Natl. Acad. Sci. U.S.A. 92:5496-5500(1995).
 RN [2]
 RP SEQUENCE OF 1-588 FROM N.A.
 RX MEDLINE=95042297; PubMed=7954395;
 RA Campbell I.G., Foulkes W.D., Senger G., Trowsdale J.,
 RA Garin-Chesa P., Rettig W.J.;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CD166 antigen precursor (activated leukocyte-cell adhesion molecule)
 DE (ALCAM) (DM-GRASP protein).
 GN ALCAM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_Taxid=10090.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NFS;
 RX MEDLINE=97353242; PubMed=9209500;
 RA Bowen M.A., Bajorath J., D'Egidio M., Whitney G.S., Palmer D.,
 RA Kobarg J., Starling G.C., Stidak A.W., Aruffo A.;
 RT "Characterization of mouse ALCAM (CD166): the CD6 binding domain is
 RT conserved in different homologs and mediates cross-species binding";
 RL Eur. J. Immunol. 27:1469-1478(1997).
 RN [2]
 RP SEQUENCE OF 227-583 FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Brain;
 RX MEDLINE=94376084; PubMed=8089660;
 RA Kanli J.P., Chang S., Kuwada J.Y.;
 RT "The molecular cloning and characterization of potential chick
 RT DM-GRASP homologs in zebrafish and mouse.";
 RL J. Neurobiol. 25:831-845(1994).
 CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN
 CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC
 CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO
 CC ACTIVATED LEUKOCYTES, AS WELL AS IN INTERACTIONS BETWEEN CELLS OF
 CC THE NERVOUS SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U95030; AAC06342.1; -;
 DR EMBL: L25274; AAA37528.1; -;
 DR HSSP: Q13740; IAC.
 DR MGD: MGI:131326; ALCAM.
 DR InterPro: IPR003599; I9.
 DR InterPro: IPR003006; I9_MHC.
 DR InterPro: IPR003600; I9_Like.
 DR Pfam: PF00047; I9; 5.
 DR SMART: SM00409; I9; 3.
 DR SMART: SM00410; I9_Like; 2.
 DR PROSITE: PS00290; I9_MHC; FALSE_NEG.
 KW Cell adhesion; Immunoglobulin domain; glycoprotein; Transmembrane;
 KW Repeat; Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 583 CD166 ANTIGEN.
 FT DOMAIN 28 527 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 528 549 POTENTIAL.
 FT DOMAIN 550 583 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 321 321 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 347 399 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 428 492 IG-LIKE C2-TYPE DOMAIN 3.
 FT DISULFID 43 113 POTENTIAL.
 FT DISULFID 157 220 POTENTIAL.
 FT DISULFID 270 313 POTENTIAL.
 FT DISULFID 354 392 POTENTIAL.
 FT DISULFID 435 485 POTENTIAL.
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 227 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 583 AA; 65161 MW; E7BAF8FCA8F9489 CRC64;
 S -> F (IN REF. 2).
 Query Match 14.5%; Score 486.5; DB 1; Length 583;
 Best Local Similarity 24.5%; Pred. No. 5,9e-25;
 Matches 152; Conservative 122; Mismatches 274; Indels 73; Gaps 25;
 5 RLVCAFLAACCCCPRAGVGEAEOPAPRLVVEVSGTALCKGSGQGNLSHVDPS 64
 10 RLVPCLLSAVALRPLGLWV-----TVNSATGDTIVMCRDLVQ-NLMEFKMY 58
 65 VHKERLIFRVGOGQS---EPGEYERLSLDGATLALTYTPDERTFLCQ-GK 119
 59 EKPDSVVFIAFSSSTKSYQYDDEYKDLTSE-NYTLISANKISDEKRFVCLVT 117
 120 RPRQERIDRYKAPKEENIQVNPGLIPVNSKEPEVATCGRGYPIQVIWYKNGR 179
 118 EDNVEAPTLVYKFKQSPKEI-VNKAPE-LETDQLKKGDICSDSYDPGNTWYRNEK 175
 180 PLKEKRRVLIQSSQVTE-SSGLYTLQSLKAQVLKEDKADQFVCLNRLPSGNMKSS 238
 176 VLPQVEEVALIRKEKEDPTQLYTSSLEYTTSDIOMPTGCVTYYPGQGTIIS 235
 239 REVTVPVFPTKRWLEV-EPVGMKEGDRVEIRCLADGNPPH---FSTKONPSTREA 294
 236 EQEIFDIYPTQETQIOLVLPKNAIKEDGNITLQCLGNGNPPEEFMFYLPQGPESIRSS 295
 295 EERTNNGVIVLEPAKHSRGVEGQGLDLMISLSPQLLNYSVDRVSPAAP- 353
 296 NTVYLTN-----VRNATGDKCSLDKRNMAAST---ITVHYL-DLSLRGEV 342
 354 ERQGSLLTTCLEASSQDLFEQMLRETEQVLERGPVLQHLDKREAGGYRCVAVSPS 413
 343 TKQIGDLPVPSCTISASRNATVYMKDNIR--LRSSP--SFSSLYHQDAGNYCETALDE 398
 414 IPGINRQVLVNAIFCPPMAFKRWKVENVNLSCASHPPTISMVNGTAS-- 471
 399 VEGLKRESLTLVEGKP-QIKMTKTPSGSKTIICVGEFPKPAIMHTTGGSVI 456
 472 -EODDPO--RYLSTLNVLTPELLETGEVCTASNDIGKNTSILFELVNLTTLPDSDN 527
 457 NQTESPYINGRYS--KIISPEENVY-LTCTAENDLETVNSLVNSAISI---PE-- 507
 528 TTGTLSTSTASPHTRANSTSTERKLPEPESRGVYIAVVICIYLAVALYELV-KKG 586
 508 -----HDEADSIDENR-EKVNDQAKIYGIYVGLLALVAGVYVYLMKRS 554
 587 KLPRBSGKEITLPPSRKSE 607
 555 KTASKHVNKDLGNMEENKLE 575
 RESULT 6
 ID C166_CARAU STANDARD; PRT; 555 AA.
 AC Q90304;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CD166 antigen homolog precursor (Neurolin) (DM-GRASP homolog).
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Cyprinodontiformes; Teleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Carassius.
 NC NCB1_Taxid=7957;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Retina;
 RA Laessing U.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RT [2]
 RP SEQUENCE OF 30-552 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Retina;
 RX MEDLINE=94299040; PubMed=8026643;
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuenkel C.A.O.;
 RT "Molecular characterization of fish neurulin: a growth-associated
 cell surface protein and member of the immunoglobulin superfamily in
 the fish retinectal system with similarities to chick protein
 DM-GRASP/SC-1/BEN";
 RT Differentiation 56:21-29(1994).
 RL
 CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
 CC NEURONS VIA HETEROPHILIC AND HOMOPHILIC INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: PRESENT ON ALL RETINAL GANGLION CELLS (RGCs)
 CC AND THEIR AXONS (IN EMBRYO). ABSENT FROM MATURE AXONS ALONG MOST
 CC OF THEIR LENGTH, BUT IS PRESENT ON NEW AND GROWING AXONS DERIVED
 CC FROM THE RGCs AT THE RETINAL MARGIN. REMAINS ON ADULT RGCs ONLY AT
 CC CELL-CELL CONTACT SITES AND IS CONTINUOUSLY FOUND IN THE RETINAL
 CC AXON TERMINAL ARBOR LAYERS OF THE ADULT TECTUM.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L25056; AAC38015.2; -
 DR HSSP: Q13740; 1KJC.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: Pf00047; Ig; 3.
 DR SMART: SM00410; Ig_Like; 3.
 DR PROSITE: PS00290; Ig_MHC; 1.
 DR Signal: Developmental protein: Cell adhesion; Immunoglobulin domain;
 KW Repeat: Glycoprotein; Transmembrane.
 KW
 FT SIGNAL 1 22
 FT CHAIN 23 555
 FT DOMAIN 23 499 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 500 520 POTENTIAL.
 FT DOMAIN 521 555 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 31 117 IG-LIKE V-TYPE DOMAIN 1.
 FT DOMAIN 147 224 IG-LIKE V-TYPE DOMAIN 2.
 FT DOMAIN 256 313 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 340 392 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 419 477 IG-LIKE C2-TYPE DOMAIN 3.
 FT DISULFID 38 110 POTENTIAL.
 FT DISULFID 154 217 POTENTIAL.
 FT DISULFID 263 306 POTENTIAL.
 FT DISULFID 426 470 POTENTIAL.
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 555 AA; 60371 MW; 5A4A8014F0DBFF68 CRC64;

Query Match 13.8%; Score 465.5; DB 1; Length 555;
 Best Local Similarity 26.5%; Pred. No. 1,8e-23;
 Matches 167; Conservative 111; Mismatches 234; Indels 119; Gaps 27;

QY 6 LVCAFLAAC-----CCCPVAGVPGEAQAPAEVVEGVSTALLKC--GLSQSGNLSH 59
 DB 7 LIGAFIAAAVFRPGSGVGVIGLGE-----TIYPCNDGTRKPPDG-LIF 50
 QY 60 VDFSVYHKKRLLIRRVGGOG-----SEPGYEQRLSLDRGATLALTYTPDERFL 115
 DB 51 TWKVKYKDDSGPDGLLVKQAQKDEATVSATDGYKSRVSLA-ANSSLLARSLDQRYFT 109
 QY 116 CQ-GKRPRSEYRIOLRYVKAPEEPNIOVNPGLGVNSKEPE-----EVAICVGRNGVPI 169
 DB 110 CMVYSTNLEETVEKWKHKSAPYIK-----NNAKLENGKLTQIGECYVENANPP 162
 QY 170 PQTWYKNGRPILKEKNRHHIOSQTVES-SGLYTLQSLTKAQLVKEKDAQFYCELYNR 228
 DB 163 ADLWKKNNQTLVDGKTLITITKTKITGLSSSTSSRLQYTKARKEVDSQFTC----- 217
 QY 229 LBSGNM-----KESREYVYVYPIEKWLEVEPYGMKKEGDRVETRLADGNPP---H 281
 DB 218 --TAKHVMGPDVYSEPESEPIHYPIPEKYSLOVVSOSPINEGDTVLKQADGNPPTSFN 275
 QY 282 FSTSKQNPSTREAEETTDNGVLYLEPARKHSGRREGOGDLDTMTLSLSEPOELLVN 341
 DB 276 FNT-----KQKATYVDKDYITLTGYTRADSGITKCSLDNDVMS-----TQFYTVS 323
 QY 342 YVSDVVSAPAPE-ROEGSLTLTCEAESSODLEFQWLREETGOVLERGPVQLDHLKRE 400
 DB 324 FL-DVSLPTFGKLVKNVGNLLVSLDKNASSEAKYTWMD--NRKLDKLP--DFSKLYTS 378
 QY 401 AGGTYCVAVSPIPLANTOLVNAVIRPPMA--FRKRWVVENNVNLSCASGHP 458
 DB 379 DADLYVCV---SIEGKISLSELTVEGIPKITSITFKRSSDGHKV--LTCAEESP 432
 QY 459 RPIISNVNNGTSEQDQDQPVLTNLVYTPPELLLETGECTASNDLGNSTILELEVN 518
 DB 433 KPVQCVSVNGTNDVSYNNKGATYKLTVPKSNLT--VSCLVTKKIGEDFKEI----- 483
 QY 519 LITLTPDSNTTGLSTVASPHTRANSTSTERKLPBPSRGVYVAVYCIILVLAVALCAV 578
 DB 484 -----SVFSQKNEDGTE-----QARVIVGIYVGLVLAALVAL 516
 QY 579 LYFLY-KKGLPGRSKQEI-TLPPSRKE 607
 DB 517 IYWIYIKTRQSGWKTEKEAGTSESKLE 547
 RESULT 7
 C166_BRARE STANDARD; PRT; 564 AA.
 ID C166_BRARE
 AC Q90460; Q90480;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE C166 antigen homolog precursor (Neurulin) (DM-GRASP homolog).
 GN C166.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Danio.
 NCBI_TaxID=7955;
 RN
 RP SEQUENCE FROM N.A.
 RP MEDLINE=94299040; PubMed=8026643;
 RA Rankl J.P., Chang S., Kuwada J.Y.;
 RT "The molecular cloning and characterization of potential chick
 RT DM-GRASP homologs in zebrafish and mouse.";
 RL J. Neurobiol. 25:831-845(1994).
 RN
 RP SEQUENCE OF 398-561 FROM N.A.
 RP MEDLINE=94299040; PubMed=8026643;
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuenkel C.A.O.;
 RT "Molecular characterization of fish neurulin: a growth-associated
 RT cell surface protein and member of the immunoglobulin superfamily in
 RT the fish retinectal system with similarities to chick protein

RT DM-GRASP/SC-1/BEN. ;
RT Differentiation 56:21-29(1994).
CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
CC NEURONS VIA HETEROFILIC AND HOMOPHILIC INTERACTIONS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L25273; AAA50024.1; -
DR EMBL: L25057; AAA50048.1; -
DR HSSP: Q13740; IKUC.
DR ZFIN: ZDB-GENE-990415-30; cd166.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003600; IG_Like.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00410; IG_Like; 3.
DR SMART: SM00406; IG_V; 1.
DR PROSITE: PS00290; IG_MHC; 1.
KV Cell adhesion: Immunoglobulin domain; Glycoprotein; Transmembrane;
KV Repeat: Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 1 25 564 CD16 ANTIGEN HOMOLOG.
FT DOMAIN 25 507 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 508 528 POTENTIAL.
FT DOMAIN 529 564 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 117 IG-LIKE V-TYPE DOMAIN 1.
FT DOMAIN 147 222 IG-LIKE V-TYPE DOMAIN 2.
FT DOMAIN 256 314 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 340 389 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 419 477 IG-LIKE C2-TYPE DOMAIN 3.
FT DISULFID 38 110 POTENTIAL.
FT DISULFID 154 217 POTENTIAL.
FT DISULFID 263 306 POTENTIAL.
FT DISULFID 426 470 POTENTIAL.
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 488 497 LFEEDKRPFG -> R (IN REF. 2).
FT SEQUENCE 564 AA; 61273 MW; ADAA/C293A607929 CRC64;
Query Match 13.0%; Score 438; DB 1; Length 564;
Best Local Similarity 26.0%; Pred. No. 1.2e-21;
Matches 161; Conservative 112; Mismatches 254; Indels 92; Gaps 26;
OY 7 VCAFLLAACCCPRVAGVGEAEAPAPLVEVEVSTALLKGLSOSGNSHVDWFSVH 66
DB 12 IAAALFAGSCLPVIYIGYET-----IEVPCNN-----GNKKPDG-LIFTKMKYAK 57
OY 67 KEKRLIFRVNGGQ-----SEPGYEDRLSLQDRGATLALTYVTPDERFLFCQ-GKRP 121
DB 58 DDGSPGDLILQAOQKDDPTVSMQDVKTRVSA-ANSSLLAOSGLTDQRFCTMVVST 116
OY 122 RSOEYRIOLRVYKAPPEENIOWNPGLGVNSKEPEEVATCGRNGYPLPOVIYTKNGRPL 181
DB 117 NLEESVVEVKKHKSASAVIK-NKYELENK-LTQLECYVESNPADILIMKNNAL 174
OY 182 KEKNRVHVSQTVES-SGLTYQSLKLAQLVKEDKDAQFYCEINLRLPSGNHMKESRE 240
DB 175 VDDGKTIITSDVTKDPTGLSTSSSSRLQYTAKEKDVASQFTC-----YAKHVTGNQ 227
OY 241 VTVP-----VFPTPEVWLEVEPVGMLKGDVREYRCLADGNPPP---HFSISKONPSTRE 293

DB 228 VSTPDTFQIRYPTKEVSLQVVSQSPIREDDVTLCQADGNPPTSFNFNI-----K 279
OY 294 ABEETDNGVLVEPARKHESGVEQGLDLMISLSEPDGLVINYSDVRVSPAP 353
DB 280 GKRVYVDKDYTLTGLTVRADSGVYKCSLLDNDVMS-----TQIVVSFL-DASLPTPGK 334
OY 354 E-ROEGSSLTJTCBAESSODLEFQWLBRENGOYVERGPVQLDLKKEAGGYRCAVASP 412
DB 335 VLKLTGELVYSLKKNASSEKVTWTD--NRKLDKLP--DFSQLRSDAGLYCDV--- 387
OY 413 SIPLNRTQLVNVAIFSPWMA--FKERKYVENVNLNLSCEASGHPRTISMVNGTA 470
DB 388 SIEIKHSFSELPTEVEGGPRITGLTKHRSNDGKHV---LTCEAGSGPKPEQVMSVNGTD 444
OY 471 SEQQDQQRVLTNLVLTPELLETGYECIASNDLGKNTSIFLELVNLTTLPPDSWTTT 530
DB 445 DETSYVNGKATYKLTLPVPSKNLT--VSCLVTKLGFDT-----KDISVPSLEEKP 496
OY 531 GLSTSTASPHTRANSTSTERKLPESRGVYIAVYICILVLAVALGYFLY-KKGLTP 589
DB 497 G-----KNEDGADQAVYGVVGLFLAALVGLIYLIYIKTKRQ 537
OY 590 CRSGKQET-TLPPSRKE 607
DB 538 SMKTGEKETGTSSESKILE 556
RESULT 8
ICCR_DROME STANDARD; PRT; 764 AA.
ID ICCR_DROME 008180;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Irregular chiasm C-roughest protein precursor (IRRC protein).
GN RST.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94102535; PubMed=7503814;
RA Ramos R.G., Igloi G.L., Lichte B., Baumann U., Maier D.,
RA Schneider T., Brandsstaetter J.H., Froehlich A., Fischbach K.-F.,
RT "The irregular chiasm C-roughest locus of Drosophila, which affects
RT axonal projections and programmed cell death, encodes a novel
RT immunoglobulin-like protein.";
RT Genes Dev. 7:2533-2547(1993).
RL -1- FUNCTION: REQUIRED FOR CORRECT AXONAL PATHWAY FORMATION IN
CC THE OPTIC LOBE AND FOR PROGRAMMED CELL DEATH IN THE DEVELOPING
CC RETINA.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: POSTEMBRYONIC EXPRESSION IS STRONG IN THE
CC DEVELOPING OPTIC LOBE AND IN THE EYE IMAGINAL DISC.
CC -1- DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN EMBRYOS. ALSO FOUND
CC IN LATE LARVAL AND PUPAL STAGES.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z21641; CAAT9756.1; -
DR EMBL: L11040; AAA16532.1; -
DR PIR: A49448; A49448.

Query Match	8.0%;	Score 270;	DB 1;	Length 764;
Best Local Similarity	20.7%;	Pred. No. 2.2e-10;		
Matches 137;	Conservative	89;	Mismatches 277;	Indels 158;
				Gaps 25

[illegible]

Db	532	--LMTVGG-----ISVVAVLTVLTL-VVVY----	555
Qy	585	KGLDPCRRSGKOEITLPPSRKSELVEVKSDKLPEMGLLGSSGDKRAPDGSEKIIDL	644
Db	556	---IKCKNRK----LP------ADVISEHOITKNGVSKLEPEDRITSNDL	597
Qy	645	R 645	
Db	598	K 598	
<hr/>			
RESULT	9		
PTK7_CHICK			
ID_PTK7_CHICK	STANDARD:	PRT:	1051 AA.
AC	Q91048:		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DE	30-MAY-2000 (Rel. 39, Last annotation update)		
DR	Tyrosine-protein kinase-like 7 precursor (kinase like protein).		
PTK7 OR KLG.			
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
CC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryonic brain;		
RX	MEDLINE=91271300; PubMed=1711213;		
RA	Chou Y.-H., Hayman M.J.:		
RT	"Characterization of a member of the immunoglobulin gene superfamily		
RT	that possibly represents an additional class of growth factor		
RT	receptor";		
RL	Proc. Natl. Acad. Sci. U.S.A. 88:4897-4901(1991).		
CC	-I- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY		
CC	THE CATALYTIC ACTIVITY OF TYROSINE KINASE.		
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN BONE MARROW, SPLEEN, BURSA,		
CC	THYMUS AND BRAIN. WEAKLY EXPRESSED IN FIBROBLASTS. ALSO EXPRESSED		
CC	IN EMBRYONIC LIVER.		
CC	-I- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.		
CC	-I- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-		
CC	PROTEIN KINASES. TRK-TYPE SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/		
CC	or send an email to license@isb-sdb.ch).		
CC	-----		
DR	EMBL; M63437; AAA48933.1; -		
CC	HSPB; P00523; ZPTRK.		
DR	InterPro: IPRO00719; Euk_pkinase.		
DR	InterPro: IPRO03006; Ig_MHC.		
DR	InterPro: IPRO03598; Ig_C2.		
DR	InterPro: IPRO03600; Ig_like.		
DR	InterPro: IPRO01245; Tyr_pkinase.		
DR	Pfam: PF00047; Ig_7.		
DR	Pfam: PF00069; pkinase; 1.		
DR	PRINTS: PR00109; TYRKINASE.		
DR	SMART: SMO0410; IG_Like; 2.		
DR	SMART: SMO0408; IGC2; 4.		
DR	SMART: SMO0219; Tyrcg; 1.		
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.		
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.		
KW	Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;		
KW	Immunoglobulin domain; Repeat.		
FT	SIGNAL	1	22
FT	CHAIN	23	1051
FT	DOMAIN	23	685
FT	TRANSMEM	686	706
FT			POTENTIAL.
FT			TYROSINE-PROTEIN KINASE-LIKE 7.
FT			EXTRACELLULAR (POTENTIAL).
FT			POTENTIAL.

```

QY 29 EGPAPELVEVEGSGTALLKCKLSQSGNLSHVD-----WF-----SVHK 67
Db 119 KQPA-SAAELIQSSSTVLRNC-----HIDGHPRTMQWFGDAPLPDGRKGYSSVS 167
QY 68 EKRTLIFVRGOGOGSEGEY-----EORTLDDRCATLALQ-----V 106
Db 168 KERLTILR---GAGPDDGLYYCARPARVSGSQNFNLNITDEFPAAVVPEDLY 224
QY 107 TPQDRITLQ--GKRPSQSY-----RIQLR---VYK--- 134
Db 225 TKNEAMPDQCPAAVPRPTQMLFEDSPITNRSKTYFANGSLITLQVKARSTGYKCI 284
QY 135 -----APEEPNIOVPLGIYVNSKEPEVATCGRNGYPIPOV 172
Db 285 HGQKGKALVYKATLRLAEIEMAFSPKVLTAQGNHV-----SCACPRGVPPTPOV 335
QY 173 IWKYK-----GRPLKEKNRHHIOSSQTVESGLTYTOSILKAQLYEKDQAFYCEL 225
Db 336 WMRNQRVPTAGRVYQEAQLVF--TSITEADAGIYTCNAKKA----- 378
QY 226 NYRLPSGNHMKESRETVVYFPRYEKWKLEVEFYGMKEDRDVEITGLADGNPPHFSIS 285
Db 379 -----GEKKQELSTIV---ATVPKVMMPKDSQLESKPGYLCTLKASIKPTVTWY 427
QY 286 KQNSTREBAEETTNDNGVLLEPARKEHSGRECOGLDIDTWTISLSEPOELVNY--- 342
Db 428 RNVYSISEDSRFEISENGTLRINNVEYDGTMYK-----VSSTPASSIEGTAVY 477
QY 343 --VSDVAVSPAAPERQ---EGSSLTITLCEASSQDLEFQWLRKE---TGQVLERGPVLQ 393
Db 478 HVLEKLEFTPPQPLQCMENKRYTSCATGREKPTIOMTKDGSLSHVSHRAGILS 537
QY 394 LHDLCRAGGRCVAVSPISPLGNTOLYLVNLFGRPMARK--ERKWKVENVNVLNIS 451
Db 538 FHKVSRSDSNYTCIASNSPOGIRKATVOLVAVY---VTRKLEPEPTTVYQGTAFMQ 593
QY 452 CEASGAPRPTISWNVNGTASEDQDQRYLSTLVNVLPELLETGECTASNDLGKNTSI 511
Db 594 CQAEQDPVPHIQWK---GMDKILDPKSLPLRIQWPNKSLV---IYDVTDSKRYICI 646
QY 512 -----LFLELVNLTITLPPDSNTTGTSTASHTANSTERKLIPEESRGV 560
Db 647 AGNSCNIKHREALFLVYVDKPAAEEDEG-----PSSHPTPKMIQT-----IGL 688

```

Query Match Best Local Similarity 7.7%; Score 259.5; DB 1; Length 1051;
Matches 149; Conservative 89; Mismatches 263; Indels 241; Gaps 31;

```

QY 29 EGPAPELVEVEGSGTALLKCKLSQSGNLSHVD-----WF-----SVHK 67
Db 119 KQPA-SAAELIQSSSTVLRNC-----HIDGHPRTMQWFGDAPLPDGRKGYSSVS 167
QY 68 EKRTLIFVRGOGOGSEGEY-----EORTLDDRCATLALQ-----V 106
Db 168 KERLTILR---GAGPDDGLYYCARPARVSGSQNFNLNITDEFPAAVVPEDLY 224
QY 107 TPQDRITLQ--GKRPSQSY-----RIQLR---VYK--- 134
Db 225 TKNEAMPDQCPAAVPRPTQMLFEDSPITNRSKTYFANGSLITLQVKARSTGYKCI 284
QY 135 -----APEEPNIOVPLGIYVNSKEPEVATCGRNGYPIPOV 172
Db 285 HGQKGKALVYKATLRLAEIEMAFSPKVLTAQGNHV-----SCACPRGVPPTPOV 335
QY 173 IWKYK-----GRPLKEKNRHHIOSSQTVESGLTYTOSILKAQLYEKDQAFYCEL 225
Db 336 WMRNQRVPTAGRVYQEAQLVF--TSITEADAGIYTCNAKKA----- 378
QY 226 NYRLPSGNHMKESRETVVYFPRYEKWKLEVEFYGMKEDRDVEITGLADGNPPHFSIS 285
Db 379 -----GEKKQELSTIV---ATVPKVMMPKDSQLESKPGYLCTLKASIKPTVTWY 427
QY 286 KQNSTREBAEETTNDNGVLLEPARKEHSGRECOGLDIDTWTISLSEPOELVNY--- 342
Db 428 RNVYSISEDSRFEISENGTLRINNVEYDGTMYK-----VSSTPASSIEGTAVY 477
QY 343 --VSDVAVSPAAPERQ---EGSSLTITLCEASSQDLEFQWLRKE---TGQVLERGPVLQ 393
Db 478 HVLEKLEFTPPQPLQCMENKRYTSCATGREKPTIOMTKDGSLSHVSHRAGILS 537
QY 394 LHDLCRAGGRCVAVSPISPLGNTOLYLVNLFGRPMARK--ERKWKVENVNVLNIS 451
Db 538 FHKVSRSDSNYTCIASNSPOGIRKATVOLVAVY---VTRKLEPEPTTVYQGTAFMQ 593
QY 452 CEASGAPRPTISWNVNGTASEDQDQRYLSTLVNVLPELLETGECTASNDLGKNTSI 511
Db 594 CQAEQDPVPHIQWK---GMDKILDPKSLPLRIQWPNKSLV---IYDVTDSKRYICI 646
QY 512 -----LFLELVNLTITLPPDSNTTGTSTASHTANSTERKLIPEESRGV 560
Db 647 AGNSCNIKHREALFLVYVDKPAAEEDEG-----PSSHPTPKMIQT-----IGL 688

```

```

QY 561 VIVAIVYICLVAVIGAVLYFLYKKGKLPGRSGKQETILPPRSKSEL----- 608
Db 689 SVGAAYAVIITL--VLG--LMFYCKR-----RRKARLKKHPEGEPEPEECNGSTLQNG 739
QY 609 --VVEYKSKLPEBEGMLLOGSS 628
Db 740 QTTAEIQEAVLFTNLGSSSGAS 761

RESULT 10
NCAL_XENLA
ID NCAL_XENLA STANDARD: PRT: 1088 AA.
AC P16170:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neural cell adhesion molecule 1, large isoform precursor (N-CAM 180)
DE [Contains: N-CAM 140].
GN NCAM1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=90098771; PubMed=2481269;
RA Kriegl P.A., Sakaguchi D.S., Kintner C.R.;
RT "Primary structure and developmental expression of a large
RT cytoplasmic domain form of Xenopus laevis neural cell adhesion
RT molecule (NCAM).";
RL Nucleic Acids Res. 17:10321-10335(1989).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURON AND IN PRESUMPTIVE NEURAL
CC TISSUE.
CC -1- DEVELOPMENTAL STAGE: THE MRNA ENCODING THIS LD-NCAM IS THE MAJOR
CC TRANSCRIPT PRESENT IN BOTH MATERNAL RNA AND IN THE EMBryo DURING
CC EARLY NEURAL DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M2596; AAA49909.1; -
CC PIR; S09600; IXYLNL.
CC HSSP; P56276; IYLR.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003006; IG_MHC.
CC InterPro; IPR003598; IG_C2.
CC Pfam; PF00041; fn3; 2.
CC Pfam; PF00047; ig; 5.
CC SMART; SM00060; FN3; 2.
CC SMART; SM00408; IGC2; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
CC Immunoglobulin domain; Alternative splicing; Signal.
CC CHAIN 1
CC SIGNAL 19
CC DOMAIN 20 1088 NEURAL CELL ADHESION MOLECULE 1, LARGE
CC ISOPFORM.
CC TRANSMEM 706 723 EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.

```

FT DOMAIN 724 1088 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 34 100 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 129 193 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 225 289 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 316 386 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 413 480 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 512 589 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 618 686 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 149 153 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 158 162 HEPARIN-BINDING (POTENTIAL).
 FT DISULFID 41 93 PROBABLE.
 FT DISULFID 136 186 PROBABLE.
 FT DISULFID 232 282 PROBABLE.
 FT DISULFID 323 379 PROBABLE.
 FT DISULFID 420 473 PROBABLE.
 FT CARBOHYD 82 89 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 804 1049 MISSING (IN ISOBORN N-CAM 140).
 SQ SEQUENCE 1088 AA; 117778 MW; 6273855B03F3E83 CRC64;

Query Match 7.6%; Score 256; DB 1; Length 1088;
 Best Local Similarity 18.1%; Pred. No. 3, 1e-09;
 Matches 155; Conservative 100; Mismatches 266; Indels 336; Gaps 35;

QY 33 PELVEVEGSTALLKCGLSQGNLSHVDFVYHKEKRTLLIFRVROGQSGEGERQRL 92
 DB PDGEISLGESEKFLC---QVSGEATDISWSPTEKLVN-----QOOL 66
 QY 93 SL---QORGATLALQVTPDERIFLC-----QGR-----RPRQERY 126
 DB SVRSDDYTLTLTYNASSODAGIKVCASNEAGESEGTVNLKIYOKLTFKNAPPPQEF 136
 QY 127 R-----IOLRVKAPPE----- 138
 DB KEGEDAVIICDVSSSIITWRHKKDVLFKKDVRVYLANNTLOIRGIKTKDEGYRC 186
 QY 139 -----PNIQVPLGIPIVNSKEPEVATCVGRNGYPIPOVITWK 176
 DB EGRILARGEINVKDIQYIVNVPPIQARQVRVATNANMASVYLSCADDFPPEISMLK 246
 QY 177 NGRPLKEKNVHIQSSQYVASSSLYLOSLIKRAQLVKEKDKAQFYELVRLPSGHHM 236
 DB KGEPIEDGEKISFNE-----QSEMTIHHKEDAEVSCIAN-----NOAG 289
 QY 237 ESREVTVPVEPYREKVMLEVPVGMLEKGRVETIRCLADGNPPPHFESISKQNFSTREAE 296
 DB EAETITLLKVAKKIKTYVENKTAIV--ELDEITITCEASDPDP--SITWRTAVRNISSE 345
 QY 297 ETTDNGVLYLEPARKK-----HSGRYEC-----OGLDIDWISLSEPOELVN 341
 DB ATTLDGHIIVKHEIRMSALTLKDIQYTDAGEYFCIASNPICVDMQAMV-----FEVQ 397
 QY 342 YVSDVRVSPAPEROEGSSLTIFCEASSODLEFOWLRETFGQVLE-----RGPV 391
 DB YAPKIR-GPVVVYVTEGKPNVITTEVFAHPRAAVTWFRD--GQLPSSNSNFKIYSGPT 454
 QY 392 ---LQLDLKRAGGGRVY-----ASVPSIPOLANTOLVN--VAIFG 429
 DB SSSLEVNDSDNDGNVNCATINTIGHSEFELIYQADTFSSPAIRKVEVYSSVWMLVFD 514
 QY 430 PP-----WMAEKERVWK-----ENM--VINLSCEAS----- 455
 DB EPDSTGVPIIKYKAEKRVIGHEKMHKYYDAKEVNAESIITYGGLPETSVMKLSAMN 574
 QY 456 ---GHPRTISMVN-----GTASE-----ODOPORV-----LST 483
 DB 575 GKGLDSTPSSQEFITTPVREPSAKLVGHLSEDSNSTIKVDLKODDGGSPIRHYLVYRA 634

QY 484 LNLVY-TPEL-----LETGVE---CYASNDIGKNTSIL--FLELVNLTLIT 523
 DB 635 LNLLEKPEKMRPPSSHHVMKALKEMNVDYEVYVAENOGKRSKALLSFTTKPATYT 694
 QY 524 PSDNTTGLSTSTASPHRANSTSTERKLPEPESGKVIVAVIYCIILAVIAGVLYFLY 583
 DB 695 ATASAGTGLGTG-----AIVGILLIVFVLLVVDVYVTCFFLN 731
 QY 584 KKGKLPK-----RSGKQETLTPPSKSELVYVSKDKLPEEKGLLOG 626
 DB 732 KCGLLMCIAVNFNCGRAGPAKAGKDIEGKAAPSKDESK--EPIYEVRE----- 778
 QY 627 SSGDKRAPDQGEKYID 643
 DB 779 ---EERTPHHDSNQIE 792
 RESULT 11
 NCAM2 MOUSE
 ID NCAM2 MOUSE STANDARD; PRT; 725 AA.
 AC P13594: 061950:
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neural cell adhesion molecule, phosphatidylinositol-linked isoform
 DE precursor (N-CAM 120) (NCAM-120).
 GN NCAM1 OR NCAM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RA MEDLINE=87246524; Pubmed=3595563;
 RA Barthele D., Santoni M.-J., Wille W., Ruppert C., Caix J.-C.,
 RA Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;
 RT Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
 RT a Mr 79,000 polypeptide without a membrane-spanning region.";
 RL EMBO J. 6:907-914(1987).
 RN [2]
 RP SEQUENCE OF 20-700 FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA MEDLINE=89251563; Pubmed=2721486;
 RA Santoni M.J., Barthele D., Vopper G., Boned A., Goridis C., Wille M.;
 RT "Differential exon usage involving an unusual splicing mechanism
 RT generates at least eight types of NCAM cDNA in mouse brain.";
 RL EMBO J. 8:385-392(1989).
 RN [3]
 RP SEQUENCE OF 642-725 FROM N.A.
 RC MEDLINE=88283628; Pubmed=3396534;
 RA Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;
 RT "Differential splicing and alternative polyadenylation generates
 RT distinct NCAM transcripts and proteins in the mouse.";
 RL EMBO J. 7:625-632(1988).
 RN [4]
 RP SEQUENCE OF 20-36.
 RC MEDLINE=86140120; Pubmed=3512556;
 RA Rougon G., Marshak D.R.;
 RT "Structural and immunological characterization of the amino-terminal
 RT domain of mammalian neural cell adhesion molecules.";
 RL J. Biol. Chem. 261:3396-3401(1986).
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
 CC NEURITES, ETC.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI anchor.
 CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
 CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y00051; CAA68263.1; -.
DR EMBL: X15049; CAA33148.1; ALT_SEQ.
DR EMBL: X07195; CAA30173.1; -.
DR PIR: A29673; IJMSNG.
DR MGD: MGI:97281; Ncam.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGC2; 5.
KW Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
KW Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 725
FT DOMAIN 34 103 NEURAL CELL ADHESION MOLECULE,
FT DOMAIN 132 196 PROSPATIDYLINOSITOL-LINKED ISOFORM.
FT DOMAIN 228 295 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 323 393 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 420 487 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 519 596 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 625 692 IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN 152 156 FIBRONECTIN TYPE-III 1.
FT DOMAIN 161 165 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 41 96 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 139 189 PROBABLE.
FT DISULFID 235 288 PROBABLE.
FT DISULFID 330 386 PROBABLE.
FT DISULFID 427 480 PROBABLE.
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 268 ERSRSVS -> DEKHFRSD (IN REF. 2).
FT CONFLICT 273 273 V -> L (IN REF. 2).
FT CONFLICT 354 355 QD -> KT (IN REF. 2).
FT CONFLICT 349 349 T -> K (IN REF. 2).
FT CONFLICT 572 572 T -> R (IN REF. 2).
FT CONFLICT 575 575 D -> V (IN REF. 2).
FT CONFLICT 589 594 MOPSSES -> SAATEF (IN REF. 2).
FT CONFLICT 600 602 PEL -> REP (IN REF. 2).
FT CONFLICT 657 657 H -> D (IN REF. 2 AND 3).
SQ SEQUENCE 725 AA; 80296 MW; C2AE8B84461C6B2F CRC64;

Query Match 7.3%; Score 246.5; DB 1; Length 725;
Best Local Similarity 20.6%; Pred No. 7.4e-09;
Matches 141; Conservative 92; Mismatches 261; Indels 189; Gaps 33;

QY 33 PELVEVEGSLALNCGLSQGNLSHVD--WFSVHKEKRTLIFFVRGOGGSEGEYEQ 90
DB 26 PGESEIVGESKFFLC---QVAGDAKDKDISWFSNGEKLS-----PNOQ 67
QY 91 RLSL---QDRGATLALTQVTPDDEKFLC--QGRPRROEYRIQIQLRYKAPREPIQNP 145
DB 68 RLSTVWNNDDSSLTLYINADAGIKCYVAEGCTOSEAVANVKIRQ-----KLMEKN 122
QY 146 LGIPVNSKEPEE-VATCGVGRNGYPIPOVIWYKNGRLPEKEKN-RVHISSQTSVESSGLYT 203
DB 123 APTPOEFKGEDEAVICDVYVSSLP-PTIIMKHKGKDVILKKDVRIVLSNNYLQIRG--- 178

QY 204 LOSILKAOLVEDKDAQFCYCELYNRLPDSGNHMKSEREVTVPEYPTKEVLWEYVGMKL 263
DB 179 : : : : : IKKTDEGTYRCE-----GRLANGE-----INKKQIVLVNPPYQAR 217
QY 264 E-----GDRVEIRCLADGNPPPHFSISKQ-NPSTREAEEB---TTNDGCVLVLEP 309
DB 218 : : : : : QSIYATATNLQOSVTLVCDADGFEPEPTMSWTMKDEPEIENEEDERSRSVSDESSEVIRN 277
QY 310 ARKESGRGECQG-----LDIDTMISLSPEOBLLVNVSDVAVSPAAREGGSSTLT 363
DB 278 VDKNDEATYVCIAENKAGEODASIHLYFAFKP---ITVE-----NOTAMELEE--QVTL 328
QY 364 TCEAESQDLEFFQWLREET-----GOYLERGPV-----LOLHDLKREAGGRCVY 408
DB 329 TCEASGDPITPITW-RTSTRNISSEEDDLOGHMVVRSHAVSSLTLSKISYRAGEWMT 387
QY 409 AS-----VPSIPLGNLTQLVNVAIFPPPMARERKRVWYKENVLNLSC 452
DB 388 ASNTIGDSQSIDLEFOYAPKIQG-----PVAVY-----TW--EGNOVAVITC 427
QY 453 EASGHPRTISWNNYNGTASBEDDQQRVLSLTVLP--ELLEET-----GYECTA 501
DB 428 EVFATPSATISWFRDQGL-----PSSNYSNIKTYNTPSASYLEVPPDSNDPGNINCTA 482
QY 502 SNDGKNTSILFELVNLTLTPDSNTTGTGSTASPHTRANSTERTKLPESRGVY 561
DB 483 VNRIGQES-----LEFLVQADTPSSP-----SIDRVEPY-----SSTNQVGFDEPATGV 529
QY 562 IVAVIVCL-----VLAIVGAVLYFLYKKKKLPCRRSGKOET 599
DB 530 PILYKAEMKSLGEESWHTFYDAKEANMGIVIMGLKDETYSDRLAALANGGLGEIM 589
QY 600 LPSPKSELVYEVRKDKLPREMG 622
DB 590 QPSEKTKQPVPELSAPKLEGOMG 612

RESULT 12
VCAL RAT STANDARD; PRT; 739 AA.
ID VCAL RAT
AC P29534;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CN NCBI_TaxID=10116;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Lung.
RX MEDLINE=92181437; PubMed=1371918;
RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,
RA Burkly L., Miyake K., Kincaid P., Lobb R.;
RT Cloning of murine and rat vascular cell adhesion molecule-1.;
RL Biochem. Biophys. Res. Commun. 183:163-169(1992).
CC -!- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTERGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED ON INFLAMED VASCULAR ENDOTHELIUM, AS
CC WELL AS ON MACROPHAGE-LIKE AND DENDRITIC CELL TYPES IN BOTH NORMAL
CC AND INFLAMED TISSUE.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M84488; AAA2332.1; .
PIR: J50675; J50675.

HSP: P19320; IYCA.
InterPro: IPR003987; ICAM_YCAM-1.

InterPro: IPR003006; I9_MHC.

InterPro: IPR003598; I9_C2.

InterPro: IPR003600; I9_Like.

InterPro: IPR003989; YCAM-1.

Pfam: PF00047; I9_5.

PRINTS: PR01472; ICAMVCAM1.

PRINTS: PR01474; YCAM1.

SMART: SM00410; IG_Like; 1.

SMART: SM00408; IGC2; 4.

Immunoglobulin domain; glycoprotein; Cell adhesion; Transmembrane;

Repeat; Signal.

STGNL 1 24 PROBABLE.

CHAIN 25 739 VASCULAR CELL ADHESION PROTEIN 1.

DOMAIN 25 698 EXTRACELLULAR (POTENTIAL).

TRANSMEM 699 720 POTENTIAL.

DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).

DOMAIN 38 91 IG-LIKE C2-TYPE DOMAIN 1.

DOMAIN 129 198 IG-LIKE C2-TYPE DOMAIN 2.

DOMAIN 237 287 IG-LIKE C2-TYPE DOMAIN 3.

DOMAIN 326 379 IG-LIKE C2-TYPE DOMAIN 4.

DOMAIN 418 496 IG-LIKE C2-TYPE DOMAIN 5.

DOMAIN 525 575 IG-LIKE C2-TYPE DOMAIN 6.

DOMAIN 612 675 IG-LIKE C2-TYPE DOMAIN 7.

CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).

SEQUENCE 739 AA; 81246 MW; 5C08E5A1A1B100C CRC64;

Query Match 7.38; Score 246.5; DB 1; Length 739;

Best Local Similarity 21.5%; Pred. No. 7.6e-09;

Matches 133; Conservative 83; Mismatches 182; Indels 221; Gaps 30;

39 EVGSTALLKCGLSQSGNLISHVDFSVHKEKRTLIFFRQGGGSEPEYEQRLSLDQRG 98

38 QIGDSMLTSCSTTCES--PSFSW--RTQIDSPLNCKVTEGAK----- 77

99 ATLALQVTPQDERIFLC-----OGKPRQSEYRIOLRVYKAPPEPNIOVN-PL--GIPV 150

78 SVLTMPVSFENEHSLTCAATNSGLERG----IQVDIYSPKXPELQFSGPLEVGRPV 133

151 NSKEPEEAVTCVGRNGYPIP--QVIWYKNGRLKEKKNVHIQSSQYVSSGLYLOSL 208

134 MKV-----CLAPDYPIIDLEIELFKGDRMLKKQ--DEVDEMAKSLTKSLVYFT-- 183

209 KAOLVEDKDAQFYCELNRIKPSGNHMKSRREV--VPFYFTEKWLVEPEVGMLEKED 266

184 ---PVEDIEKALVCRAKLYIDQTDIPKEREVRELQYTSKNTNEISHVSTRHGEA 240

267 RVEIRCLADGNPPPHSISKONPSTREAEETNDNGV-----LVLEPAKHEHSR 317

241 AVNTTCASEGLPAPPEIFMSK-----LDNGVLIQLSGNATLTILAMRMEDSGI 288

318 YECGGIDL---DTMISLSEPEOLLVNYSDVRVPAAP--EQGESSLTINCEASSQD 372

289 YVCGVNLVGRDKTEVLELVQEKRFYVD-----ISGSSVAADVGDVVLITCAAVGCS 342

373 LEFQWLREET---GOVLENG----- 389

343 PSFSW-RTQIDSPLNCKVTEGAKRTSLTISPVGEDEHSLCTVTCORRLEKTIQVEVY 401

QY 390 -----PY-----IQLDHDKRE----- 400

DB 402 SFPEDEPIETISGLVGRPTVNCVPPYPPHLEIELLKGTSLINKFLREIGTKSL 461

QY 401 -----AGGYRCVA-----SVPSPLGNRIQVNVAVIFGPPMAFKE 437

DB 462 ETKSLEMTFIPFAEDTKALVCLAKLHSSQMESEPKOROSTQRLYYNVA---P---KE 513

QY 438 RKVAV-----KENMVLNLSCESGHPRTISWN---VNGTASEODODPQRLVSTLVIV 488

DB 514 PTIVSPSPVPEEGSPNLTSSSGFTPTKILMSROLKNEQLPLSQN-----TILSFMA 568

QY 489 TPPELLTGV-ECTASNDLG 506

DB 569 T-KMEDSGIYCEGINEAG 586

RESULT 13

NCAL_RAT NCAL_RAT STANDARD; PRT; 858 AA.

AC P13596:

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)

DE (NCAM-140).

DE NCAM1 OR NCAM.

OS Rattus norvegicus (rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxId=10116;

OX [1]

SEQUENCE FROM N.A.

RC TISSUE=Brain;

RC MEDLINE=88059265; PubMed=3680385;

RC Small S.J., Shull G.E., Santoni M.-J., Akeson R.;

RT Identification of a cDNA clone that contains the complete coding

RT sequence for a 140-kD rat NCAM polypeptide.";

RT J. Cell Biol. 105:2335-2345(1987).

RT [2]

RN SEQUENCE OF 355-364 FROM N.A.

RX MEDLINE=90166485; PubMed=2483093;

RX Small S.J., Haines S.L., Akeson R.A.;

RT Polypeptide variation in an N-CAM extracellular immunoglobulin-like

RT fold is developmentally regulated through alternative splicing.";

RT Neuron 1:1007-1017(1988).

CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN

CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF

CC NEURITES, ETC.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF

CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.

CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: X06564; CAA29809.1; .

CC EMBL: M32611; AAA4679.1; .

CC PIR: S00846; IJRTNC.

CC InterPro: IPR003961; FN_III.

CC InterPro: IPR003006; I9_MHC.

CC InterPro: IPR003598; I9_C2.

CC Pfam: PF00041; fn3; 2.

CC Pfam: PF00047; I9_5.

CC SMART: SM00060; FN3; 2.

DR	EMBL: M71714; AAA39911.1; -	
DR	EMBL: J04054; AAA39899.1; -	
DR	EMBL: J04055; AAA39912.1; -	
DR	HSPG: 007954; 1CR8	
DR	MED: MGI:96257; Hspg2	
DR	InterPro: IPR000561; EGF-like.	
DR	InterPro: IPR000742; EGF_2.	
DR	InterPro: IPR001438; EGF_11.	
DR	InterPro: IPR003006; 19_MHC.	
DR	InterPro: IPR003598; 19_C2.	
DR	InterPro: IPR002172; LDL_recept-A.	
DR	InterPro: IPR002043; Laminin_B.	
DR	InterPro: IPR000094; Laminin_Egf.	
DR	InterPro: IPR001791; Laminin_G.	
DR	InterPro: IPR000082; SEA.	
DR	Pfam: PF00047; 19; 15.	
DR	Pfam: PF00052; Laminin_B; 3.	
DR	Pfam: PF00053; Laminin_EGF; 8.	
DR	Pfam: PF00054; Laminin_G; 3.	
DR	Pfam: PF00057; ldl_recept_a; 4.	
DR	Pfam: PF01390; SEA_1	
DR	PRINTS: PR00101; EGFBL00D.	
DR	ProDom: PD003031; Laminin_B; 3.	
DR	SMART: SM00180; EGF_Lam; 7.	
DR	SMART: SM00001; EGF_Like; 6.	
DR	SMART: SM00408; IGC2; 14.	
DR	SMART: SM00192; LDLa; 4.	
DR	SMART: SM00281; Lamb; 3.	
DR	SMART: SM00282; Lamg; 3.	
DR	SMART: SM00200; SEA; 1.	
DR	PROSITE: PS00022; EGF_1; 8.	
DR	PROSITE: PS01186; EGF_2; 5.	
DR	PROSITE: PS01248; LAMININ_TYPE_EGF; 11.	
DR	PROSITE: PS50025; LAM_G_DOMAIN; 3.	
DR	PROSITE: PSS0209; LDLRa_2; 4.	
DR	PROSITE: PSS0068; LDLRa_2; 4.	
DR	PROSITE: PSS0024; SEA_1.	
KW	Signal; Basement membrane; Proteoglycan; Repeat; glycoprotein;	
KW	Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;	
KW	Extracellular matrix; EGF-like domain.	
FT	SIGNAL	1
FT	CHAIN	22 3707
FT	DOMAIN	80 194
FT	DOMAIN	195 234
FT	DOMAIN	281 319
FT	DOMAIN	320 359
FT	DOMAIN	360 403
FT	DOMAIN	404 504
FT	DOMAIN	521 530
FT	DOMAIN	531 730
FT	DOMAIN	731 763
FT	DOMAIN	764 813
FT	DOMAIN	814 871
FT	DOMAIN	879 923
FT	DOMAIN	924 933
FT	DOMAIN	934 1125
FT	DOMAIN	1126 1158
FT	DOMAIN	1159 1208
FT	DOMAIN	1209 1265
FT	DOMAIN	1275 1324
FT	DOMAIN	1325 1334
FT	DOMAIN	1335 1529
FT	DOMAIN	1530 1562
FT	DOMAIN	1563 1612
FT	DOMAIN	1613 1670
FT	DOMAIN	1677 1771
FT	DOMAIN	1772 1863
FT	DOMAIN	1866 1954
FT	DOMAIN	1955 2049
FT	DOMAIN	2050 2148
FT	DOMAIN	2149 2244
FT	DOMAIN	2245 2343

FT	DOMAIN	2344	2532	IG-LIKE C2-TYPE DOMAIN 9.
FT	DOMAIN	2437	2526	IG-LIKE C2-TYPE DOMAIN 10.
FT	DOMAIN	2533	2619	IG-LIKE C2-TYPE DOMAIN 11.
FT	DOMAIN	2620	2770	IG-LIKE C2-TYPE DOMAIN 12.
FT	DOMAIN	2721	2809	IG-LIKE C2-TYPE DOMAIN 13.
FT	DOMAIN	2810	2885	IG-LIKE C2-TYPE DOMAIN 14.
FT	DOMAIN	2960	2980	IG-LIKE C2-TYPE DOMAIN 15.
FT	DOMAIN	2984	3162	EGF-LIKE.
FT	DOMAIN	3163	3241	LAMININ G-LIKE 1.
FT	DOMAIN	3245	3435	LAMININ G-LIKE 2.
FT	DOMAIN	3518	3705	LAMININ G-LIKE 3.
FT	SIZE	65	67	HEPARAN SULFATE (POTENTIAL).
FT	SIZE	71	73	HEPARAN SULFATE (POTENTIAL).
FT	SIZE	76	78	HEPARAN SULFATE (POTENTIAL).
FT	SIZE	3615	3617	MEDIANE MOTOR NEURON ATTACHMENT (POTENTIAL).
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	234	BY SIMILARITY.
FT	DISULFID	285	297	BY SIMILARITY.
FT	DISULFID	292	310	BY SIMILARITY.
FT	DISULFID	304	319	BY SIMILARITY.
FT	DISULFID	325	337	BY SIMILARITY.
FT	DISULFID	332	350	BY SIMILARITY.
FT	DISULFID	344	359	BY SIMILARITY.
FT	DISULFID	368	381	BY SIMILARITY.
FT	DISULFID	375	384	BY SIMILARITY.
FT	DISULFID	388	403	BY SIMILARITY.
FT	DISULFID	428	479	BY SIMILARITY.
FT	DISULFID	764	773	BY SIMILARITY.
FT	DISULFID	766	780	BY SIMILARITY.
FT	DISULFID	783	792	BY SIMILARITY.
FT	DISULFID	795	811	BY SIMILARITY.
FT	DISULFID	814	829	BY SIMILARITY.
FT	DISULFID	816	839	BY SIMILARITY.
FT	DISULFID	842	851	BY SIMILARITY.
FT	DISULFID	854	869	BY SIMILARITY.
FT	DISULFID	1159	1168	BY SIMILARITY.
FT	DISULFID	1161	1175	BY SIMILARITY.
FT	DISULFID	1178	1187	BY SIMILARITY.
FT	DISULFID	1190	1206	BY SIMILARITY.
FT	DISULFID	1209	1224	BY SIMILARITY.
FT	DISULFID	1211	1234	BY SIMILARITY.
FT	DISULFID	1237	1246	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1275	1287	BY SIMILARITY.
FT	DISULFID	1295	1304	BY SIMILARITY.
FT	DISULFID	1297	1293	BY SIMILARITY.
FT	DISULFID	1307	1322	BY SIMILARITY.
FT	DISULFID	1363	1372	BY SIMILARITY.
FT	DISULFID	1563	1579	BY SIMILARITY.
FT	DISULFID	1582	1591	BY SIMILARITY.
FT	DISULFID	1594	1610	BY SIMILARITY.
FT	DISULFID	1613	1628	BY SIMILARITY.
FT	DISULFID	1615	1638	BY SIMILARITY.
FT	DISULFID	1641	1650	BY SIMILARITY.
FT	DISULFID	1653	1668	BY SIMILARITY.
FT	DISULFID	1792	1839	BY SIMILARITY.
FT	DISULFID	1886	1932	BY SIMILARITY.
FT	DISULFID	1976	2021	BY SIMILARITY.
FT	DISULFID	2073	2118	BY SIMILARITY.
FT	DISULFID	2170	2215	BY SIMILARITY.
FT	DISULFID	2268	2313	BY SIMILARITY.
FT	DISULFID	2365	2413	BY SIMILARITY.
FT	DISULFID	2456	2506	BY SIMILARITY.
FT	DISULFID	2554	2599	BY SIMILARITY.
FT	DISULFID	2641	2686	BY SIMILARITY.
FT	DISULFID	2831	2876	BY SIMILARITY.

Query Match	7.3%;	Score 244.5;	DB 1;	Length 3707;
Best Local Similarity	22.3%;	Pred. No. 9.5e-08;		
Matches 113; Conservative	79;	Mismatches 190;	Indels 125;	Gaps 26;

```

OY 17 CCRVAGVP-----GEAEQAPAEVLEVEY-----GSTALLKCKLSQSGNLSHYDMF 63
DB 1653 CAGYEGADNVQGGKCPPTRESLEVOIHPSRSVPQGGPHSLRC---QVSGSPPHFYFW 1709
OY 64 SVHKERKTLIFRVROGQSGSEGEYEQRLSLQDRGATLALQVTPQDERIFLCOGKR-PR 122
DB 1710 S-----REDRPRPSSAQOR-----HQSSELPSPVQSPDAGVYICTRNLIH 1752
OY 123 SOEYRIQLRVYKAPPEPNIQVNPGLGVNKSKEPEVA-----TCVGRNCGYPIQVITWY 175
DB 1753 TSSSRKELLVAPSK-----PIWYVEEQRQSQVSRPPADYTFICTAKSKSPATLVMT 1806
OY 176 K--NGRPLKEEKNRVHIQSSQVSSGLYTLQSLAQLVKEDKQAGY--CELN--YRLP 230
DB 1807 RLHNGK-----LPSRAMDFNGLITIRNV-----QPSDAGVYCTGSMNPRMD 1848
OY 231 SGN---HMKESREVTVPV--FYPTKEVYLEVEPVGMLKESGDRVEITCLADGNPPPHF-- 282
DB 1849 QGTATLHVQVSGSTAPVASIHPQ---LYVQP-----GQQAEPKSCATGNPTMLEWI 1899
OY 283 -STSKNPSTRAEEETNDNGVILPARKHSGRYEGOGD-----LDTMISL--S 333
DB 1900 GGFSGGLPAKQI--H-----NGILRLAIPRSDQGLCLRALSSAGQVYARAMQVHGG 1953
OY 334 EPBELLVNVSDVRSVPAPEKSSGLTTCABESSQDLFPQWLREETGYLER----G 389
DB 1954 GPR-----VQVSPERTVHSGRTVRLYCRAGVPSASITW--RKEGSLPRHOAHG 2003
OY 390 PVLQHLDLKREAGGGRVAVSPISIFGLNFTQLVNV-----AIFGPPMAFKERKRV 441
DB 2004 SRLRLHMSVADSGEYVRAN--NNIDQETSIMISVSPSTNSPPAPADRIESSSR 2062
OY 442 VKENNVNLNCSCEASGCHPPTSMVNV 468
DB 2063 VAEQGTLDLNCVPGHAAQYTWKRG 2089

RESULT 15
NCAL_BOVIN STANDARD: PRT: 853 AA.
ID NCAL_BOVIN P31836:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140).
GN NCAM OR NCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Kumlantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain cortex;
RX MEDLINE=89378239; PubMed=2776887;
RA LIPKIN V.M., Khramtsov N.V., Andreeva S.G., Moshnyakov M.V.,
RA Petukhova G.V., Rakitina T.V., Reshenko E.A., Ishchenko K.A.,
RA Murzeva S.F., Chernova M.N., Dranytsyna S.M.;
RA "calmodulin-independent bovine brain adenylate cyclase. Amino acid
RT sequence and nucleotide sequence of the corresponding cDNA.";
RU FEBS Lett. 254:69-73(1989).
RL [2]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
RN [3]
RP IDENTIFICATION AS N-CAM.
RX MEDLINE=9211748; PubMed=1765159;
RA Premont R.T.;
```

```

RT "A bovine brain cDNA purported to encode calmodulin-insensitive
RT adenylate cyclase has extensive identity with neural cell adhesion
RT molecules (N-CAMs).";
RL FEBS Lett. 295:230-231(1991).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A CALMODULIN-
CC INDEPENDENT ADENYLATE CYCLASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL: X16451; CA934470.1; -.
DR PIR: A32976; IJBONC.
DR HSP: P40189; IBDU.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR Pfam: PF00041; Fn3; 2.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IG2; 5.
DR Cell adhesion; Glycoprotein; Transmembrane; Repeat.
KW Immunoglobulin domain; Alternative splicing; Signal.
FT STGNL 1
FT CHAIN 20 853
FT DOMAIN 20 719
FT TRANSMEM 737
FT DOMAIN 738 853
FT DOMAIN 34 103
FT DOMAIN 132 196
FT DOMAIN 228 293
FT DOMAIN 321 401
FT DOMAIN 428 495
FT DOMAIN 527 604
FT DOMAIN 633 700
FT DOMAIN 152 156
FT DOMAIN 161 165
FT DOMAIN 161 165
FT DISULFID 139 189
FT DISULFID 235 286
FT DISULFID 328 394
FT DISULFID 435 488
FT CARBOHYD 222 222
FT CARBOHYD 314 314
FT CARBOHYD 346 346
FT CARBOHYD 432 432
FT CARBOHYD 458 458
FT CARBOHYD 487 487
SQ SEQUENCE 853 AA; 93893 MW; E12FD49231A7A368 CRC64;

Query Match 7.2%; Score 242.5; DB 1; Length 853;
Best Local Similarity 20.8%; Pred. No. 1.7e-08;
Matches 146; Conservative 93; Mismatches 245; Indels 217; Gaps 34;

OY 33 PELVEVSGSTALLKCGLSQSGNLSHVD--WFSYKKEKRTLIFRVROGQSGQSPGEYEQ 90
DB 26 PSQGEISVGSKFFLC---QVAGAKDKDISWSPGKELI-----PNDQ 67
```


QY 91 RLST---QDRGATLALQVTPDERIFLC--QGRPRSQEYRIQLRYKAPBEPNIQVNP 145
 Db 68 RISVVMNDSSSTLTIYANIDAGIYKCVTAEDGTSEATVNVKIFQ-----KLMFKN 122
 QY 146 LGIFVNSKEPEE-VATCVGRNGYPIPOVITWKGRPLKEKN-RVHIQSQIYESSGLYT 203
 Db 123 APTEPQEFREGEDAVICDVVSSLP-PTIIRKHGRDVIILKDVDFIYLTNNYLOIRG--- 178
 QY 204 LOSILKQALVKEDDAQFYCELVNRLPSGNMKESREVTVPVPEYPERKWLVEPEYGM 263
 Db 179 -----IKKIDEGTYRCE-----GRILARGE-----INFKDIQVIVNPPVTVQAR 217
 QY 264 E-----GDRVEINCLADGNPPHFSISKONPSTREAEET---TNDNGVLVLEPAR 311
 Db 218 OSIYANATANIGQSVTLVCNAGEPPEPVSWTKGBQIENEDKYLFSDDSSSLTIRKVD 277
 QY 312 KEHSGRYECOG-----LDLDTMISLSEPOELLVNVSDVRYSPAAPERQSGSSLTLC 365
 Db 278 KNEBAEVYVCIENKAGBQDASIHILKYAPKPK---ITYVE---NOTAMELEE--QVTLTC 328
 QY 366 EABE-----SODLEFQWLREETGOVLERGPLYQLH-----DKREA 401
 Db 329 EASGDPIPSITWTSTRNISSEKASWTREPKQETLDGHNVRSHARVSSLTILKSIQYTD 388
 QY 402 GGGYRCYAS-----VPSIPGLNRTOLVNVAIFGPPMMAFKERKVVYKEN 445
 Db 389 AGEVYCTASNTIGODSQSMYLEVOYAPKLOG-----PYAVY-----TW--EG 428
 QY 446 MVLNLSCASGHPPTISMNVNGTASEQDODPQRVLSTLNVLP--ELLETT----- 495
 Db 429 NOVWITCEVFAYPASATISMFRDQQL-----PSSNYSNIKIYNTPSASYLEVTPDSEND 483
 QY 496 -GVECTASNDLGKNTSLTLEIVNLTTLPDSSNTTGLSTSTASPHTRANSTSTERKLE 554
 Db 484 GNTNCTAVNKRIGES---LEFVLVQADTPSSP-----STDQVEPY---SSTAQVQFDE 530
 QY 555 PESRGVIVAV-----IVCIIVL-----AVLGAVLYF 581
 Db 531 PEATGGVPIILKYKAEMWRAEYVHWSKWDYAKESMEGIYTIYGLKPEYIAVRLAAL-- 588
 QY 582 LYKKGKLPCCRSGKQETTLPPSRKSELVYEVKSKDKLPEEMG 622
 Db 589 ---NGK-----GIGETISAASEFKTQVREPAPKLEGGMG 620

Search completed: June 28, 2002, 10:50:55
 Job time: 260 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:45:50 ; Search time 34.31 Seconds

(without alignments)
3257,206 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGPRIVCAFLAACCPCPR.....SSGDKRAPDQGEKIDLRH 646

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:.*
2: SP bacteria:.*
3: SP fungi:.*
4: SP human:.*
5: SP invertebrate:.*
6: SP mammal:.*
7: SP mhc:.*
8: SP organelle:.*
9: SP phage:.*
10: SP plant:.*
11: SP rodent:.*
12: SP virus:.*
13: SP vertebrate:.*
14: SP unclassified:.*
15: SP viirus:.*
16: SP bacteriap:.*
17: SP archaeap:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3359	99.9	646	4 Q9BRD9	Q9BRD9 homo sapien
2	3356	99.8	646	4 Q95812	Q95812 homo sapien
3	2642	78.6	648	11 Q9EPF1	Q9EPF1 mus musculu
4	2587	76.9	648	11 Q9EPF2	Q9EPF2 raltus norv
5	2409	71.6	606	11 Q9ESF7	Q9ESF7 mus musculu
6	2354	70.0	606	11 Q9ESF8	Q9ESF8 raltus norv
7	1104	32.8	626	13 Q98922	Q98922 gallus galli
8	1025	30.5	626	13 Q90880	Q90880 gallus galli
9	936	27.8	584	13 Q98921	Q98921 gallus galli
10	863	25.7	584	13 Q90989	Q90989 gallus galli
11	747	22.2	504	13 Q98923	Q98923 gallus galli
12	645.5	19.2	628	6 Q9WZ08	Q9WZ08 bos taurus
13	618	18.4	622	11 Q9JKB2	Q9JKB2 mus musculu
14	614	18.3	622	11 Q9R069	Q9R069 mus musculu
15	613	18.2	622	11 Q9ES55	Q9ES55 mus musculu
16	613	18.2	624	11 Q9ES56	Q9ES56 raltus norv

17	606	18.0	650	11 Q99K86	Q99K86 mus musculu
18	499.5	14.9	583	11 Q35112	Q35112 raltus norv
19	491.5	14.6	583	6 Q9BHL3	Q9BHL3 bos taurus
20	485	14.4	521	6 Q46634	Q46634 canis faml
21	465	13.8	521	6 Q46651	Q46651 oryctolagus
22	407	12.1	88	11 Q9JHC2	Q9JHC2 mus musculu
23	362	10.8	86	6 Q19129	Q19129 bos taurus
24	305	9.1	605	4 Q96J84	Q96J84 homo sapien
25	281	8.4	2673	4 Q96SC3	Q96SC3 homo sapien
26	281	8.4	5198	5 Q76518	Q76518 caenorhabdi
27	280	8.3	5636	4 Q96RW7	Q96RW7 homo sapien
28	269	8.0	1252	4 Q96DM3	Q96DM3 homo sapien
29	268.5	8.0	538	6 Q29123	Q29123 sus scrofa
30	264.5	7.9	538	6 Q28939	Q28939 sus scrofa
31	260	7.7	1034	4 Q96M60	Q96M60 homo sapien
32	254	7.6	1406	5 Q9GPE7	Q9GPE7 drosophila
33	253	7.5	779	4 Q96JG0	Q96JG0 homo sapien
34	248.5	7.4	846	13 Q57577	Q57577 cynops pyrr
35	248.5	7.4	1100	13 Q57576	Q57576 cynops pyrr
36	247.5	7.4	4370	4 Q9H3V5	Q9H3V5 homo sapien
37	247	7.3	1612	11 Q89026	Q89026 mus musculu
38	246.5	7.3	1651	11 Q53005	Q53005 raltus norv
39	245.5	7.3	739	11 Q63669	Q63669 raltus norv
40	245.5	7.3	1675	13 Q98SM4	Q98SM4 brachydanio
41	244.5	7.3	823	5 Q9VQ10	Q9VQ10 drosophila
42	242	7.2	739	6 Q28260	Q28260 canis faml
43	240.5	7.2	939	5 Q967X6	Q967X6 drosophila
44	240	7.1	1482	5 Q9V4Y0	Q9V4Y0 drosophila
45	237	7.0	725	13 Q73633	Q73633 xenopus lae

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	646 AA.
ID	Q9BRD9	Q9BRD9		
AC	Q9BRD9			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	SIMILAR TO MELANOMA ADHESION MOLECULE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RA	Scrausberg R.;			
RL	Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC006329; AAH06329.1; -			
DR	InterPro: IPR003598; Ig_C2.			
DR	InterPro: IPR003600; Ig_Like.			
DR	InterPro: IPR003006; Ig_MHC.			
DR	Pfam: PF00047; Ig_5.			
DR	SMART: SM00409; ITC_5.			
DR	SMART: SM00410; ITC2_3.			
DR	SMART: SM00410; ITC2_3.			
KW	Immunoglobulin domain.			
SO	SEQUENCE 646 AA; 71608 MW; 1P2CG96864B16635 CRC64;			
Query Match	99.9%;	Score 3359;	DB 4;	Length 646;
Best Local Similarity	99.8%;	Pred. No. 1.5e-267;		
Matches 645;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1 MGPRIVCAFLAACCPCPRVAGVGEAEPAPDELVEVGSFLKCGLSQSGNLSHY 60			
DB	1 MGPRIVCAFLAACCPCPRVAGVGEAEPAPDELVEVGSFLKCGLSQSGNLSHY 60			
OY	61 DMFSVHKETLTLFRVQGGSGSEGEYEQRLSLQDRGATLALQVTPDERIFLQGRK 120			

```
Db 61 DWFVHKEKRTLLIFRVROGOGSEBGEYEQRLSLQDGAIALTLQVTPDDEIRIFLCQCKR 120
Oy 121 PRSOEYRIQLRVYKABEENIOWNPLGIPVNSKEPEEVATCVGRNGYPIPOVIMYKNGRP 180
Db 121 PRSOEYRIQLRVYKABEENIOWNPLGIPVNSKEPEEVATCVGRNGYPIPOVIMYKNGRP 180
Oy 181 LKEEKNRVHIQSSQTVESGLTYTLOSILKAQVLEKEDKQAFYCELYRLPISGNHMKESRE 240
Db 181 LKEEKNRVHIQSSQTVESGLTYTLOSILKAQVLEKEDKQAFYCELYRLPISGNHMKESRE 240
Oy 241 VTVPEVPTKEVWLEVEPVGMLEKGDVEIRCLADGNPPPHFSISKQNPSTREAEETT 300
Db 241 VTVPEVPTKEVWLEVEPVGMLEKGDVEIRCLADGNPPPHFSISKQNPSTREAEETT 300
Oy 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRVSPAAPERQEGSS 360
Db 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRVSPAAPERQEGSS 360
Oy 361 LITLCEAESSQDLFEQWMLREBTGOVLERGVYQLDHLDKREAGGRCVAVSPISIGLNRT 420
Db 361 LITLCEAESSQDLFEQWMLREBTGOVLERGVYQLDHLDKREAGGRCVAVSPISIGLNRT 420
Oy 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISMVNGTASQDQDQPRV 480
Db 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISMVNGTASQDQDQPRV 480
Oy 481 LSTLVNLTVPBELLETVGECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Db 481 LSTLVNLTVPBELLETVGECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Oy 541 TRANSTSTERKLPPEESRGVIVAVICILVAVLGAVLFLFKKKKLCRCRSKQOETTL 600
Db 541 TRANSTSTERKLPPEESRGVIVAVICILVAVLGAVLFLFKKKKLCRCRSKQOETTL 600
Oy 601 PPSRKSELVVEYKSDKLPPEMGLLOGSSGDKRAPDQGEKYIDL RH 646
Db 601 PPSRKSELVVEYKSDKLPPEMGLLOGSSGDKRAPDQGEKYIDL RH 646

RESULT 2
Oy 095812 PRELIMINARY: PRT: 646 AA.
AC 095812;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN P1H12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gul L., Chang L., Browne P.V., Heibel R.P.;
RT "P1H12 from human umbilical vein endothelial cells."
RL EMBL; AF089868; AAD17799.1; -.
DR InterPro: IPR003598; I9_C2.
DR InterPro: IPR003600; I9_Like.
DR InterPro: IPR003006; I9_MHC.
DR Pfam: PF00047; I9_5.
DR SMART: SM00408; I9_C2; 2.
DR SMART: SM00410; I9_Like; 2.
KW Immunoglobulin domain.
SQ SEQUENCE 646 AA; 71665 MW; 1B5FB8AD930738E CRC64;

Query Match 99.8%; Score 3356; DB 4; Length 646;
Best Local Similarity 99.8%; Pred. No. 2,7e-267;
Matches 645; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MGJPLRVCAFLAACCCPRVAGVGEADQAPAEIVVEVSGTALLKGLSQSGCNLSHV 60
```

```
Db 61 DWFVHKEKRTLLIFRVROGOGSEBGEYEQRLSLQDGAIALTLQVTPDDEIRIFLCQCKR 120
Oy 61 DWFVHKEKRTLLIFRVROGOGSEBGEYEQRLSLQDGAIALTLQVTPDDEIRIFLCQCKR 120
Db 61 DWFVHKEKRTLLIFRVROGOGSEBGEYEQRLSLQDGAIALTLQVTPDDEIRIFLCQCKR 120
Oy 121 PRSOEYRIQLRVYKABEENIOWNPLGIPVNSKEPEEVATCVGRNGYPIPOVIMYKNGRP 180
Db 121 PRSOEYRIQLRVYKABEENIOWNPLGIPVNSKEPEEVATCVGRNGYPIPOVIMYKNGRP 180
Oy 181 LKEEKNRVHIQSSQTVESGLTYTLOSILKAQVLEKEDKQAFYCELYRLPISGNHMKESRE 240
Db 181 LKEEKNRVHIQSSQTVESGLTYTLOSILKAQVLEKEDKQAFYCELYRLPISGNHMKESRE 240
Oy 241 VTVPEVPTKEVWLEVEPVGMLEKGDVEIRCLADGNPPPHFSISKQNPSTREAEETT 300
Db 241 VTVPEVPTKEVWLEVEPVGMLEKGDVEIRCLADGNPPPHFSISKQNPSTREAEETT 300
Oy 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRVSPAAPERQEGSS 360
Db 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRVSPAAPERQEGSS 360
Oy 361 LITLCEAESSQDLFEQWMLREBTGOVLERGVYQLDHLDKREAGGRCVAVSPISIGLNRT 420
Db 361 LITLCEAESSQDLFEQWMLREBTGOVLERGVYQLDHLDKREAGGRCVAVSPISIGLNRT 420
Oy 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISMVNGTASQDQDQPRV 480
Db 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISMVNGTASQDQDQPRV 480
Oy 481 LSTLVNLTVPBELLETVGECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Db 481 LSTLVNLTVPBELLETVGECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Oy 541 TRANSTSTERKLPPEESRGVIVAVICILVAVLGAVLFLFKKKKLCRCRSKQOETTL 600
Db 541 TRANSTSTERKLPPEESRGVIVAVICILVAVLGAVLFLFKKKKLCRCRSKQOETTL 600
Oy 601 PPSRKSELVVEYKSDKLPPEMGLLOGSSGDKRAPDQGEKYIDL RH 646
Db 601 PPSRKSELVVEYKSDKLPPEMGLLOGSSGDKRAPDQGEKYIDL RH 646

RESULT 3
Oy 09EP1 PRELIMINARY: PRT: 648 AA.
AC 09EP1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE L-GICERIN/MUC18.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Taira E., Okumura S., Miki N.;
RT "mouse gicerin/MUC18."
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB035508; BAB16050.1; -.
DR WGI: WGI:1933966; Mcam.
DR InterPro: IPR003599; I9.
DR InterPro: IPR003598; I9_C2.
DR InterPro: IPR003600; I9_Like.
DR InterPro: IPR003006; I9_MHC.
DR Pfam: PF00047; I9_5.
DR SMART: SM00409; I9_4.
DR SMART: SM00408; I9_C2; 4.
DR SMART: SM00410; I9_Like; 5.
```

Immunoglobulin domain.
KW SEQUENCE 648 AA; 71511 MW; EA63AB35B8C28FD CRC64;

Query Match 78.6%; Score 2642; DB 11; Length 648;
Best Local Similarity 76.4%; Pred. No. 1.4e-208;
Matches 495; Conservative 73; Mismatches 78; Indels 2; Gaps 1;

```
QY 1 MGPLRYCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGLSQSGNLS 58
DB 1 MGPLRYCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGLSQSGNLS 60
QY 59 HDWFSVHKERTLFFRVHOGGSEPEYEQRLSLDGRATLALQVTPDERIFLCOG 118
DB 61 QVWFLIHKERQILFFRVHOGGSEPEYEQRLSLDGRATLALSHYTPDERIFLCOG 120
QY 119 KRPRSGEYRIOLRVYKAPERNIOVNPGLIPVNSKEPEEVATCGRNXPPIQVIWYKNG 178
DB 121 KRPRLODHYVLOVYKAPERNIOVNPGLIPVNSKEPEEVATCGRNXPPIQVIWYKNS 180
QY 179 RPLKEKNRVHIOSSQTVESSGLYTLQSLILKALVKEKDAQFYCELNYRLPSGNHAKES 238
DB 181 LPLQEEENRVHIOSSQTVESSGLYTLQSLILKALVKEKDAQFYCELNYRLPSGNHAKES 240
QY 239 REVTVPVFYPTKRWLEVEPVGMLEKGDVREIRCLADGNPPHPSISKONPSTREAEEET 298
DB 241 KEVTVPVFYPTKRWLEVEPVGMLEKGDVREIRCLADGNPPHPSISKONPSTREAEEET 300
QY 299 TNDNGVYLVEPARKHSGRVECOGLDPTMISLSEPOELLVNVSVSVRSPAPEROEG 358
DB 301 TDENGILSLPEAKHSGRVECOGLDPTMISLSEPOELLVNVSVSVRSPAPEROEG 360
QY 359 SSLITLCEASSODLEFQWMLREETGOVLENGPVLOLHLKREAGGYRCVAVSPSTIGLN 418
DB 361 ESLITLCEASSODLEFQWMLREETGOVLENGPVLOLHLKREAGGYRCVAVSPSTIGLN 420
QY 419 RTQLVNVALFGPPMAKFKRWYKENVNLNLSCEASGHPPTISMVNGTASQDODDPO 478
DB 421 RTQLVNVALFGPPMAKFKRWYKENVNLNLSCEASGHPPTISMVNGTASQDODDPO 480
QY 479 RVLSTLNVLPPELLETGECTASNDLGKNTSILFELVNLTTLPDSNTTGTSTSTAS 538
DB 481 TVSTLNVLPPELLETGECTASNDLGKNTSILFELVNLTTLPDSNTTGTSTSTAS 540
QY 539 PHTRANSTSTERKLPESRGVYIYAVYICILVAVLGAVALYELKKKGLPCRRSGKOEI 598
DB 541 PHTRANSTSTERKLPESRGVYIYAVYICILVAVLGAVALYELKKKGLPCRRSGKOEI 600
QY 599 TLPPSKSELVYEVKSDKLPPEMGLLOGSGDKRAPGOGGEXYIDLRLH 646
DB 601 TLPPSKSELVYEVKSDKLPPEMGLLOGSGDKRAPGOGGEXYIDLRLH 648
```

RESULT 4
Q9EPF2 PRELIMINARY; PRT; 648 AA.
AC Q9EPF2; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE L-GICERIN/MUC18.
GN L-GICERIN/MUC18.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Taira E., Okumura S., Miki N.;
RT "Functional analysis of rat gicerin/MUC18.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AB035506; BAB16048.1; -

DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00409; Ig; 4.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; Ig_Like; 5.
KW Immunoglobulin domain.
SEQUENCE 648 AA; 71372 MW; F8C334E0F1938A9F CRC64;

Query Match 76.9%; Score 2587; DB 11; Length 648;
Best Local Similarity 74.5%; Pred. No. 4.8e-204;
Matches 483; Conservative 80; Mismatches 83; Indels 2; Gaps 1;

```
QY 1 MGPLRYCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGLSQSGNLS 58
DB 1 MGPLRYCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGLSQSGNLS 60
QY 59 HDWFSVHKERTLFFRVHOGGSEPEYEQRLSLDGRATLALQVTPDERIFLCOG 118
DB 61 QVWFLIHKERQILFFRVHOGGSEPEYEQRLSLDGRATLALSHYTPDERIFLCOG 120
QY 119 KRPRSGEYRIOLRVYKAPERNIOVNPGLIPVNSKEPEEVATCGRNXPPIQVIWYKNG 178
DB 121 KRPRLODHYVLOVYKAPERNIOVNPGLIPVNSKEPEEVATCGRNXPPIQVIWYKNS 180
QY 179 RPLKEKNRVHIOSSQTVESSGLYTLQSLILKALVKEKDAQFYCELNYRLPSGNHAKES 238
DB 181 RPLKEKNRVHIOSSQTVESSGLYTLQSLILKALVKEKDAQFYCELNYRLPSGNHAKES 240
QY 239 REVTVPVFYPTKRWLEVEPVGMLEKGDVREIRCLADGNPPHPSISKONPSTREAEEET 298
DB 241 KEVTVPVFYPTKRWLEVEPVGMLEKGDVREIRCLADGNPPHPSISKONPSTREAEEET 300
QY 299 TNDNGVYLVEPARKHSGRVECOGLDPTMISLSEPOELLVNVSVSVRSPAPEROEG 358
DB 301 TDENGILSLPEAKHSGRVECOGLDPTMISLSEPOELLVNVSVSVRSPAPEROEG 360
QY 359 SSLITLCEASSODLEFQWMLREETGOVLENGPVLOLHLKREAGGYRCVAVSPSTIGLN 418
DB 361 DSLITLCEASSODLEFQWMLREETGOVLENGPVLOLHLKREAGGYRCVAVSPSTIGLN 420
QY 419 RTQLVNVALFGPPMAKFKRWYKENVNLNLSCEASGHPPTISMVNGTASQDODDPO 478
DB 421 RTRRVSIGIFGSPMAKFKRWYKENVNLNLSCEASGHPPTISMVNGTASQDODDPO 480
QY 479 RVLSTLNVLPPELLETGECTASNDLGKNTSILFELVNLTTLPDSNTTGTSTSTAS 538
DB 481 TVSTLNVLPPELLETGECTASNDLGKNTSILFELVNLTTLPDSNTTGTSTSTAS 540
QY 539 PHTRANSTSTERKLPESRGVYIYAVYICILVAVLGAVALYELKKKGLPCRRSGKOEI 598
DB 541 PHTRANSTSTERKLPESRGVYIYAVYICILVAVLGAVALYELKKKGLPCRRSGKOEI 600
QY 599 TLPPSKSELVYEVKSDKLPPEMGLLOGSGDKRAPGOGGEXYIDLRLH 646
DB 601 TLPPSKSELVYEVKSDKLPPEMGLLOGSGDKRAPGOGGEXYIDLRLH 648
```

RESULT 5
Q9ES57 PRELIMINARY; PRT; 606 AA.
AC Q9ES57; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE S-GICERIN/MUC18.
GN MUS MUSCULUS (Mouse).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
EMBL; AB035506; BAB16048.1; -

OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RA Taira E., Okumura S., Miki N.;
RT "mouse s-glycerin/MUC18."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB035509; BAB16051.1;
DR MGI; MGI:1933966; Mcam.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 5.
DR SMART; SM00409; Ig; 4.
DR SMART; SM00408; IgC2; 4.
DR SMART; SM00410; Ig_Like; 5.
DR Immunoglobulin domain.
SQ SEQUENCE 606 AA; 66867 MW; 8F3CBAC0A5F29B10 CRC64;

Query Match 71.6%; Score 2409; DB 11; Length 606;
Best Local Similarity 75.0%; Pred. No. 1.9e-189;
Matches 450; Conservative 72; Mismatches 76; Indels 2; Gaps 1;

QY 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKCGLSOSOGNLS 58
DB 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKCGLSOSOGNLS 58
QY 59 HDMFVSVHKERKRLIFRVNQGQSEPEGEYERLSTLDRGATLALTOYTPODESTIFLCOG 118
DB 61 QVWFILHKEKRLIFRVNQGQSEPEGEYERLSTLDRGATLALTOYTPODESTIFLCOG 120
QY 119 KRPSOEYRIOLRVYKAPPEPNIOVNPGLIPYNSKEPEEVAATCVGRNGYPIPOVYWKNG 178
DB 121 KRRRLDQHVLEQLQVAPPEPTIOANVGIHVDLQELKQVCELVYRPGSNHMKES 238
QY 179 RPLKEKNRVNHIOSQTVESGLYTLQSLILKAQLYKEDKQAFCELVYRPGSNHMKES 238
DB 181 RPLQEEENRVNHIOSQTVESGLYTLQSLILKAQLYKEDKQAFCELVYRPGSNHMKES 240
QY 239 REVTVAVFPTKRVWLEVEPVGMKKEGDRVETIRCLADGNPPHPSISKNPSTREAEEET 298
DB 241 KEVTVAVFPTKRVWLEVEPVGMKKEGDRVETIRCLADGNPPHPSISKNPSTREAEEET 300
QY 299 TNDNGVLYEPARKHSGRYECQGLDIDTMTISLSEPOELLVNVYSDVRSAPAPROEG 358
DB 301 TNDNGVLYEPARKHSGRYECQGLDIDTMTISLSEPOELLVNVYSDVRSAPAPROEG 360
QY 359 SSLTLTCEAESODLEPQMLREETGOVLENGPVQLQHLDKREAGGGRVAVSPISPIGLN 418
DB 361 SSLTLTCEAESODLEPQMLREETGOVLENGPVQLQHLDKREAGGGRVAVSPISPIGLN 420
QY 419 RTQLVNVAIFGPVMAKERKRVKEMVNLNLSCEASGHPRPITSNVNGTASBEDDOPQ 478
DB 421 RTQLVNVAIFGPVMAKERKRVKEMVNLNLSCEASGHPRPITSNVNGTASBEDDOPQ 480
QY 479 RVLSTLVNLYTPLELLETGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTTGLSTSTAS 538
DB 481 TVSTLVNLYTPLELLETGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTTGLSTSTAS 540
QY 539 PHTRANSTSTERKLPEPESRGVIVAVIYCIYLAVALGAVLELKKKKLRCRSGKOET 598
DB 541 PHTRANSTSTERKLPEPESRGVIVAVIYCIYLAVALGAVLELKKKKLRCRSGKOET 600

RESULT 6
Q9ESS8 PRELIMINARY; PRT; 606 AA.
AC Q9ESS8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE S-GICERIN/MUC18.
OS S-GICERIN/MUC18.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;
RA Taira E., Okumura S., Miki N.;
RT "rat s-glycerin/MUC18."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB035507; BAB16049.1;
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 5.
DR SMART; SM00409; Ig; 4.
DR SMART; SM00408; IgC2; 3.
DR SMART; SM00410; Ig_Like; 5.
DR Immunoglobulin domain.
SQ SEQUENCE 606 AA; 66727 MW; 6D0892443E8B8C47 CRC64;

Query Match 70.0%; Score 2354; DB 11; Length 606;
Best Local Similarity 73.0%; Pred. No. 6.3e-185;
Matches 438; Conservative 79; Mismatches 81; Indels 2; Gaps 1;

QY 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKCGLSOSOGNLS 58
DB 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKCGLSOSOGNLS 58
QY 59 HDMFVSVHKERKRLIFRVNQGQSEPEGEYERLSTLDRGATLALTOYTPODESTIFLCOG 118
DB 61 QVWFILHKEKRLIFRVNQGQSEPEGEYERLSTLDRGATLALTOYTPODESTIFLCOG 120
QY 119 KRPSOEYRIOLRVYKAPPEPNIOVNPGLIPYNSKEPEEVAATCVGRNGYPIPOVYWKNG 178
DB 121 KRRRLDQHVLEQLQVAPPEPTIOANVGIHVDLQELKQVCELVYRPGSNHMKES 238
QY 179 RPLKEKNRVNHIOSQTVESGLYTLQSLILKAQLYKEDKQAFCELVYRPGSNHMKES 238
DB 181 RPLQEEENRVNHIOSQTVESGLYTLQSLILKAQLYKEDKQAFCELVYRPGSNHMKES 240
QY 239 REVTVAVFPTKRVWLEVEPVGMKKEGDRVETIRCLADGNPPHPSISKNPSTREAEEET 298
DB 241 KEVTVAVFPTKRVWLEVEPVGMKKEGDRVETIRCLADGNPPHPSISKNPSTREAEEET 300
QY 299 TNDNGVLYEPARKHSGRYECQGLDIDTMTISLSEPOELLVNVYSDVRSAPAPROEG 358
DB 301 TNDNGVLYEPARKHSGRYECQGLDIDTMTISLSEPOELLVNVYSDVRSAPAPROEG 360
QY 359 SSLTLTCEAESODLEPQMLREETGOVLENGPVQLQHLDKREAGGGRVAVSPISPIGLN 418
DB 361 SSLTLTCEAESODLEPQMLREETGOVLENGPVQLQHLDKREAGGGRVAVSPISPIGLN 420
QY 419 RTQLVNVAIFGPVMAKERKRVKEMVNLNLSCEASGHPRPITSNVNGTASBEDDOPQ 478
DB 421 RTQLVNVAIFGPVMAKERKRVKEMVNLNLSCEASGHPRPITSNVNGTASBEDDOPQ 480
QY 479 RVLSTLVNLYTPLELLETGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTTGLSTSTAS 538
DB 481 TVSTLVNLYTPLELLETGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTTGLSTSTAS 540
QY 539 PHTRANSTSTERKLPEPESRGVIVAVIYCIYLAVALGAVLELKKKKLRCRSGKOET 598
DB 541 PHTRANSTSTERKLPEPESRGVIVAVIYCIYLAVALGAVLELKKKKLRCRSGKOET 600

RESULT 7
Q98922 PRELIMINARY; PRT; 626 AA.
ID Q98922

AC 098922:
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMCAM PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H.B190V+; TISSUE=THYMUS;
RX MEDLINE=97133433; PubMed=8978830;
RA Valilio O., Dunon D., Aissi F., Dangy J.P., McNaghy K.M., Imhof B.A.;
RT "HEMCAM, an adhesion molecule expressed by c-kit+ progenitors.";
RL J. Cell Biol. 135:1655-1668(1996).
DR EMBL: Y08855; CA70080.1; -
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 3.
KW Immunoglobulin domain; Signal.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 626 POTENTIAL.
SQ SEQUENCE 626 AA; 69104 MW; 27F8F2F47044E163 CRC64;

Query Match 32.8%; Score 1104; DB 13; Length 626;
Best Local Similarity 36.3%; Pred. No. 4.3e-82;
Matches 251; Conservative 116; Mismatches 221; Indels 68; Gaps 16;

QY 8 CAFILAACCCPRVAVGAEAPAPELVEVGSATLKGCLG-SQSGNLSHYDMVSH 66
DB 22 CPELCCLLCC-----GAAGRLVEMPAVLEVEISTARLECSFISPGNASTSIEMFYVN 76
QY 67 KEKRTLFRVRGOG-----QSEPEYBQRLSLDRGATLALQVTPDE-RIFLCQ 117
DB 77 RRP-----LRSGEAVRHHAGVRIDETEYSERLSVGEDKA-LSISKVTRQDNARTICQ 129
QY 118 --GKRPRSGEYRIQLRYKAPBEENIOVNPGLGIVNSKEPEVATCGVRCNGYPIPOVIWY 175
DB 130 VGADSGVGESRTELYTKIPAPPEITPNSAGIPAOQNDMLKLAQCTSENSFSPNITWY 189
QY 176 KNGRPLEKEKNRVHIQSSQVSSGLTYLQSLKAOLYKEDKDAQFCENLYRPLSPSNHM 235
DB 190 KNGEPLQEDKTKILITLVRESNGLYTVSTLFSKYTREDRNSLFFCTYHYWLOGQMRT 249
QY 236 KESREYVVPVFPPEKWLVEVP-VGMKEGDVREIRCLADGNPPPHFSISKQ--NPSTR 292
DB 250 KDSRVNVTYVFPTEHVELRVATNAGIVKEGDVYKLVCDADGNPAVPSFRRRLGDSWQ 309
QY 293 EAEETTNNDNGVLEVPAPRKHSRGYECOGILDTMTSLSEPOELLVNVYSVRV--SP 350
DB 310 DMTSLATNDGVLMLHNVSKSSGLYRCQTLDDMTQHEGD-VELVYNTIEGVQVMEP 368
QY 351 AAPRQSGSSLTITCEASSODLEFQWLREFTGVLERGPVLOLHDKRKRAAGGYRCVAS 410
DB 369 SSP-LHEGDSVRLSCTASHSPVKLDYQW-RDARGKVAEGNQLLTNTLFTFTSSNFSGRVK 426
QY 411 VPSITGLNRQOLVNVATFGPPMAAFKFKKRWKVENMYLNTSCASGHPRTTISMNANGTA 470
DB 427 ARSPVGLSSQVAVVAKGPRIVATSAAPLYRQDEVINTCKAIAFPQPSFMSINGT 486
QY 471 SEQDDQPRVLTNLVATPELLETGECTASNDLGKNTSILFLELVNLTTLPPDSITT 530
DB 487 HEY-MENQHASNLTIVASHDLRAGAMCYSNALGVSE-----KHQLDQKPS----- 535
QY 531 GLSTSTASPHTRANSTSTERKLPEPESRGVYIAVYICILVAVLGAVALFTLKKGKLP 590
DB 536 -----ESKGIIIVAIIVICLVAVLGSIIYFLKKKKKISC 570

QY 591 RRSKGQETLTPPSRKSELVYEVKSDKLPEEMGLQSSGDKRAPGQGEKYIDLRLH 646
DB 571 GRSGQDITKPEARKDNVKEVSKDLSEBAGILQGANKEKRPADOSEKYIDLRLN 626

RESULT 8
Q09880 PRELIMINARY; PRT; 626 AA.

AC Q09880;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE L-GICERIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEGHORN; TISSUE=GIZARD;
RX MEDLINE=96081930; PubMed=749368;
RA Taira E., Nagino T., Taniura H., Takaha N., Kim C., Kuo C., Li B.,
RA Higuchi H., Miki N.;
RT "Expression and functional analysis of a novel isoform of glicerin, an
immunoglobulin superfamily cell adhesion molecule.";
RL J. Biol. Chem. 270:28681-28687(1995).
DR EMBL: D49849; BAA08648.1; -
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 2.
KW Immunoglobulin domain.
SQ SEQUENCE 626 AA; 69074 MW; C4791EBC2EC55133 CRC64;

Query Match 30.5%; Score 1025; DB 13; Length 626;
Best Local Similarity 37.4%; Pred. No. 1.3e-75;
Matches 248; Conservative 114; Mismatches 219; Indels 82; Gaps 19;

QY 8 CAFILAACCCPRVAVGAEAPAPELVEVGSATLKGCL-----SQSGNLSHYDM 62
DB 22 CPELCCLLCC-----GAAGR-----EYIMPSSA-----GSGRHSQAQVQLLR-PW 63
QY 63 FSVHKEKRTLIFR--VRGOG-----QSEPEYBQRLSLDRGATLALQVTPDE 111
DB 64 ECLLHLRVVLCQPRPLRSGEAVRHHAGVRIDETEYSERLSVGEDKA-LSISKVTRQDN 122
QY 112 -RIFLCQ--GKRPRSGEYRIQLRYKAPBEENIOVNPGLGIVNSKEPEVATCGVRCNGY 168
DB 123 ARFTIQVGADSGVGESRTELYTKIPAPPEITPNSAGIPAOQNDMLKLAQCTSENSFP 182
QY 169 IPOVIYKNGRPLEKEKNRVHIQSSQVSSGLTYLQSLKAOLYKEDKDAQFCENLYR 228
DB 183 SPVITWYKNGEPLQEDKTKILITLVRESNGLYTVSTLFSKYTREDRNSLFFCTYHYW 242
QY 229 LPSGNIMKSSREYTVFPPEKWLVEVP-VGMKEGDVREIRCLADGNPPPHFSISKQ 287
DB 243 LOGQMRTKDSRVNVTYVFPTEHVELRVATNAGIVKEGDVYKLVCDADGNPAVPSFRR 302
QY 288 --NPSTREAEETTNNDNGVLEVPAPRKHSRGYECOGILDTMTSLSEPOELLVNVYS 345
DB 303 ELGDSQDWTSLADNRDGVLMHNVSKSSGLYRCQTLDDMTQHEGD-VELVYNTIEG 361
QY 346 VRY--SPAAPERQSGSSLTITCEASSODLEFQWLREFTGVLERGPVLOLHDKRKRA 403
DB 362 VQVKMEPSSP-LHEGDSVRLSCTASHSPVKLDYQW-RDARGKVAEGNQLLTNTLFTFT 419
QY 404 GYRCVASVSTIGLNRQOLVNVATFGPPMAAFKFKKRWKVENMYLNTSCASGHPRTTIS 463

Db 420 NESCRVAKARSVPGLEQSKOYAAVAVKGRPRIVATISAPLYVRODEVINLTCKAIAFPQPSFH 479
QY 464 WNVGTASSEDQDOPQRYLSTLNLVLTPELLETGETECTASNDLCKNSIIFLELVNLTTLT 523
Db 480 WSVNGTTHXY-MENQHMASNLTVRVSHDLRAGAMCRVSNALGVSE----KHQLDQK 533
QY 524 PDSVTTTGLSTASTPHTRANSTSTERKLPEPESRGVIVAVIICILVAVLAVITFLY 583
Db 534 PS-----ESKGIITVAIIVICILVAVLAVLSIIFLH 563
QY 584 KKGRLPCRSKQEBITLPPSRKSELVYVKSDDLPEEMGLQSSGDKRAPDQGEKYID 643
Db 564 KKGITSCGRSGKODITKPEARKDKNVVEKSDKLSEAGLQGANAEKRSPPADQSEKYID 623
QY 644 LRH 646
Db 624 LRN 626

RESULT 9
Q98921 PRELIMINARY; PRT: 584 AA.
ID Q98921
AC Q98921:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEMCAM PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=H.B190V+; TISSUE=THYMUS;
RA MEDLINE=9713433; PubMed=8978830;
RX Valiño O., Dunon D., Alsai F., Dangy J.P., McNaghy K.M., Imhof B.A.,
RT "HEMCAW, an adhesion molecule expressed by c-klt+ progenitors.";
RL J. Cell Biol. 135:1655-1668(1996).
DR EMBL: Y08856; CAZ70081.1; -
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 3.
KM Immunoglobulin domain; Signal.
FT SIGNAL 1 35
FT CHAIN 36 584 POTENTIAL.
SQ SEQUENCE 584 AA; 64422 MW; 5972D946604AF6F3 CRC64;

Query Match 27.8%; Score 936; DB 13; Length 584;
Best Local Similarity 35.9%; Pred. No. 2, 5e-68;
Matches 218; Conservative 109; Mismatches 213; Indels 68; Gaps 16;

QY 8 CAFLLAACCCCRVAGVGEAEPAPLEVEVSGTALLKGLS-QSOGNLSHWDFSVH 66
Db 22 CFFLLILCC-----GAGRLEVMYPAVLEVGSTARLECSFSIPGNASFSIEMFVN 76
QY 67 KKRRLIFRRVROGQ-----QSEPEGEORLSLDGRATLALTOYTPDE-RIFLQ 117
Db 77 RRP-~~-----~~LRSGEAVRHNASGVRIDETEYSERLSVGEKKA-LTISKYTRQDNARTFICQ 129
QY 118 --GKRPRSEYRIOLRVYKAPPEEPINQVPIGIPVNSKEPEEVAICVGRNGPIPIQVLY 175
Db 130 VGADSGGVSESTRELYTKIPAPPETTPNSAGIPNOSNDMLKIAOCTSENPSNITWY 189
QY 176 KNGRPLKEEKNRVHIQSSQTVESGLYTLQSLTKALQVLEKDKDAOFYELNRLPSGNHM 235
Db 190 KNGEPILQGEDEKTKITITLTVRSNGIYTVSLFSKYTREDNLSLFCHVHWLQGGQKRT 249
QY 236 KESREVTVPVEYPTKEVWLEVP-VGMLEKGRVETIRCLADGNPPHPSISKQ--NPSTR 292

Db 250 KDSPRVNTVTFEPTBEHVELRVATNAGIYEGDDVLTVDADGNPAVPSPFRRELGSWQ 309
QY 293 EMBEETDNGVLYLEPAKREHSGRYECOGDLDPMISLSPEOPALVNVSDVAV--SP 350
Db 310 DMTSLADTNDGVLMHNHNSKSSSGLKRCOTLDLDMTOHEGD-VELVNVYIEGVQVKNRP 368
QY 351 AAPEROEGSLTLTCEAESODLEFOWLREETGOYLERGPVLQHLDKREAGGGRVAVS 410
Db 369 SSP-LHEGDSVRLSCAHSPVKLDYQW-RDARGRYAEGNQLLTNLNFTESINSNCRVYK 426
QY 411 VSIPELANTOLVNVVAIFPPMAKEREKRYWKEMVNLNLSAAGHPRPISNVNCTA 470
Db 427 ARSVPGLEQSKOYAAVAVKGRPRIVATISAPLYVRODEVINLTCKAIAFPQPSFH 486
QY 471 SEDQDOPQRYLSTLNLVLTPELLETGETECTASNDLCKNSIIFLELVNLTTPDPSNTT 530
Db 487 HEY-MENQHMASNLTVRVSHDLRAGAMCRVSNALGVSE----KHQLDQKPS----- 535
QY 531 GLSTASPTSTRANSTSTERKLPEPESRGVIVAVIICILVAVLAVITFLYKKGKLP 590
Db 536 -----ESKGIITVAIIVICILVAVLAVLSIIFLHKRGKISC 570
QY 591 RRSKQEI 598
Db 571 GRSKQDI 578

RESULT 10
Q90989 PRELIMINARY; PRT: 584 AA.
ID Q90989
AC Q90989:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE S-GICERIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=LEGHORN; TISSUE=GIZZARD;
RX MEDLINE=94213753; PubMed=8161457;
RA Taira E., Tachana N., Taniura H., Kim C., Miki N.;
RT "Molecular cloning and functional expression of glicerol, a novel cell
adhesion molecule that binds to neurite outgrowth factor.";
RL Neuron 12:861-872(1994).
RN [2]
RP SEQUENCE FROM N. A.
RC STRAIN=LEGHORN; TISSUE=GIZZARD;
RX MEDLINE=96081930; PubMed=749388;
RA Taira E., Nagino T., Taniura H., Takana N., Kim C., Kuo C., Li B.,
RA Higuchi H., Miki N.;
RT "Expression and functional analysis of a novel isoform of glicerol, an
immunoglobulin superfamily cell adhesion molecule.";
RL J. Biol. Chem. 270:28681-28687(1995).
DR EMBL: D38559; BAA07563.1; -
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 2.
KM Immunoglobulin domain.
SQ SEQUENCE 584 AA; 64378 MW; 876E0C3E920BA92F CRC64;

Query Match 25.7%; Score 863; DB 13; Length 584;
Best Local Similarity 35.1%; Pred. No. 2, 6e-62;
Matches 216; Conservative 107; Mismatches 210; Indels 82; Gaps 19;


```
QY 8 CAFLAACCCPRVAGVGEAEOPAPELVEVNGSTALLKCGL-----SOSQGNLSHVDV 62
FT 11
DB 22 CPEFLCLCC-----GAAGRP-----EYFMPSA-----GSGDRHSAGVOLLH-PW 63
QY 63 FSVHKEKRTLIFR---VRQGG-----QSEGEYERLSLODRATLALQVTPQDE 111
DB 64 ECLHLNRVYLCQPRPRSGEAVRHNASGVRIDETSEYERLSVGEDNA-LSISVYTRQDN 122
QY 112 -RIFLCQ--GKRPRSOEYRIOLRVYKAPPEPNIOVNPGLIPVNSKEPEEAVTCGRNGY 168
DB 123 ARTICQVAGDSQGVESRTELTYTKIPAPPEITPNSAGIPAGSNDMLKIAQCTSENSFP 162
QY 169 IPQVITWKNRGPKEKNRVHIOSSQTVESGGLTYLQSLKADLYEKDKDAQFVCELNVR 228
DB 183 SPNTITWKNRGPKEKILTTLVRESNGLYTVSTLFSKYTRDRNSLFHCTVHYW 242
QY 229 LPSNHHKESSEYVVPFPEKRWLEVER-VGMLKEGDRVEIRCLADGNPPPHFSISKQ 287
DB 243 LQGMRTKDSPRVNTVTFEYPTHEVELRVAATNAGIVKEGDVKTLCVCDADGNPAPVFSFR 302
QY 288 --NPSTREAEETNDNGVLYLEPARKHSGRYECQGLDLDTMISLSSEPOELLVNVSD 345
DB 303 ELGSMQDMSTSLADTNDGVLYLHNVKSSSGILRCQTLDDMTQHEGD-VELVYNTIEG 361
QY 346 VRV--SPAAPERQSGSLTLTCEAESSQDLEFQWLRBETQVLEGRPVQLHDLKREAG 403
DB 362 VQVMEKPSRP-LHGGDSVRLSCTASHSPVKLDYQW-RDARGRKVAEGNQLLTLNLFETS 419
QY 404 GYRCVAVSPSTPGKNTQOLVNAIFGPPMAFKERKRVKEMNVLNLSCEASGHPRTIS 463
DB 420 NESCVRKARSPGLEQSKOVAVAVKGRPRIVAISAPLYRQDEVINTCKAIAFPSPSFH 479
QY 464 MNVNGTASEDQDQORLYSTLNLVLPPELLTETGECTASNDLGKNTSIFLEVLNLTTLT 523
DB 480 MSVAGTTHEY-MENQHMASNLTIVRSHDLRAGMCRVSNALGSE-----KHQLDLQK 533
QY 524 PDSMTTGLSTSTASPTHRANSTSTERKLPEPESRGVYVAVYVLCILVLAVALYFLY 563
DB 534 PS-----ESKGIIIVAIIVCLVAVLSSIIYFLH 563
QY 584 KKGKLPGRSGKORI 598
DB 564 KKGKISCGRSGKODI 578

RESULT 11
ID 098923 PRELIMINARY; PRT; 504 AA.
AC 098923;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 11
RP SEQUENCE FROM N.A.
RA STRAIN=H.B190V+; TISSUE=THYMUS;
RX MEDLINE=9713433; PubMed=8978830;
RT Valin O., Dunon D., Aissi F., Dandy J.P., McNagay K.M., Imhof B.A.;
RT "HMCAM, an adhesion molecule expressed by c-kit+ progenitors.";
RL J. Cell Biol. 135:1655-1668(1996).
DR EMBL, Y08854; CAZ70079.1;
DR InterPro: IPR003596; Ig_E2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 2.
KW Immunoglobulin domain; Signal.
```

```
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 504 POTENTIAL.
SQ SEQUENCE 504 AA; 55540 MW; AD4EAB94ED4F02E1 CRC64;
```

Query Match 22.2%; Score 747; DB 13; Length 504;
Best Local Similarity 35.9%; Pred. No. 7,1e-53;
Matches 172; Conservative 91; Mismatches 184; Indels 32; Gaps 13;

```
QY 8 CAFLAACCCPRVAGVGEAEOPAPELVEVNGSTALLKCGLS-QSQGNLSHVDVSH 66
DB 22 CPEFLCLCC-----GAAGRLVYMPAYLEVEIGSTARLECSFSPGNASFTSIEMFYVN 76
QY 67 KEKRTLIFVRQGG-----QSEPEYERLSLODRATLALQVTPQDE-RIFLCQ 117
DB 77 RRP-----LRGEAVRHNASGVRIDETSEYERLSVGEDEKA-LSISKVTRQDNARTFICQ 129
QY 118 --GKRPRSOEYRIOLRVYKAPPEPNIOVNPGLIPVNSKEPEEAVTCGRNGYPIPVITW 175
DB 130 VGADSQGVESRTELTYTKIPAPPEITPNSAGIPAGSNDMLKIAQCTSENSFPSPNITW 189
QY 176 KNGRPLKEKNRVHIOSSQTVESGGLTYLQSLKADLYEKDKDAQFVCELNVRPSGNH 235
DB 190 KNGEPLQEDTKILTTLVRESNGLYTVSTLFSKYTRDRNSLFHCTVHYWLOGQWRT 249
QY 236 KESREYTVFVFPTEKRWLEVER-VGMLKEGDRVEIRCLADGNPPPHFSISKQ--NPSTR 292
DB 250 KDSPRVNTVTFEYPTHEVELRVAATNAGIVKEGDVKTLCVCDADGNPAPVFSFRRELQDSMQ 309
QY 293 EAEETTNDNGVLYLEPARKHSGRYECQGLDLDTMISLSSEPOELLVNVSDVRV--SP 350
DB 310 DMTSLADTNDGVLYLHNVKSSSGILRCQTLDDMTQHEGD-VELVYNTIEGVYKMEP 368
QY 351 AAPERQSGSLTLTCEAESSQDLEFQWLRBETQVLEGRPVQLHDLKREAGGRCVYAS 410
DB 369 SSP-LHGGDSVRLSCTASHSPVKLDYQW-RDARGRVAREGNQLLTLNLFETSSNFSRVK 426
QY 411 VPSIPGLNTQOLVNAIFGPPMAFKERKRVKEMNVLNLSCEASGHPRTISMVNVT 469
DB 427 ARSVPLEQSKOVAVAVKGRPRIVAISAPLYRQDEVINTCKAIAFPSPSFHMSINGT 485
```

```
RESULT 12
ID 09M208 PRELIMINARY; PRT; 628 AA.
AC 09M208;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LUTHERAN GLYCOPROTEIN.
GN LU.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20363335; PubMed=10908043;
RA Boado R.J., Li J.Y., Pardridge W.M.;
RT "Selective Lutheran glycoprotein gene expression at the blood-brain
RT barrier in normal brain and in human brain tumors.";
RL J. Cereb. Blood Flow Metab. 20:1096-1102(2000).
DR EMBL, AF270512; AAF81749.2;
SQ SEQUENCE 628 AA; 68002 MW; 2128B1F1B73E6A93 CRC64;
```

Query Match 19.2%; Score 645.5; DB 6; Length 628;

Best Local Similarity 30.3%; Pred. No. 2.2e-44;
Matches 205; Conservative 106; Mismatches 263; Indels 103; Gaps 28;

OY	4	PRVCAFLIAACCCCRVAVGPEAPQPAPELVEVGSTALLKGLSGSQGNLSHV62
Db	14	PRLLVLALLA-----PRESKAENVLPPLVEVNRGSVTLDOS-PLGTHDYEMLEW 66
OY	63	FSVHK-----EKRTLIFVRQ--GGQSEPGEGEQRSLSDRGATLALLOYTPOD 110
Db	67	FIVDRSGARHRLASAEIRGSELROKEILNSRSP----YQLDSQGR---LVLPDAQVGD 119
OY	111	ERILECCK--RRPSQEFRIOLRYKKAPDEPNIOVNLGIPIVANSKEEFVATCVGRGY 168
Db	120	ERDVIYCVAKAGAACTAETALKYFAKPEAPEVSPNKGILTVMDDFAQETATCSSRGNP 179
OY	169	IPOVIVYKNGRPLEK--EKNRVHIQSQQT--ESSGLTLOSIIKAOLYKDKCAOFCYL 225
Db	180	AQOLIMWRNGRPPLAVPLEVNSEGWTITRTYREASGLSLSTLYLRHKRDREASFICSV 239
OY	226	NYRLPSGNHM-KESRETVPVFTYEKY--WLEVEPY--GMLEKGRVETRICLAGDNPP 280
Db	240	HYYLPAGOHGLDPSPSSLTLYHPTEHLFWLGSGSTAEGWVRGDSVOLLCOGDSDPTP 299
OY	281	HESI-----SKONPSTRAEAEETTNDNGVLVLEPARKEHSGREYECQGLDPTMISL-LSEP 335
Db	300	EYTFPMLODKREDYLK-----TSLEGMLTERQNQSGTYGCVRYPEDVPEDAELSKT 353
OY	336	QELLNVYSDVVRSPAAPEROE--GSSLTTCAESSDOLEFQWLREETGVLERGPVL 392
Db	354	LELRAYADLSLEL--AGEELSLEPLHNSTYTVCARGLPPTYLYMTDSA--FMGEDPTL 409
OY	393	QLHDLRKAGGCGYCVAVSPIPLANTOLYNVAFIPPMMAFERK-----PWVENMV 447
Db	410	SLHSTYEDSACTYCEAMRPIPLISTRSRLVYGSTPELKAKEIOPRAEGSWTEDEV 469
OY	448	LNLSCASGHRPTISMN-VNGTASEDDQPOR-VLSTLVNLYTPELLETVECTASNDL 505
Db	470	-TLICYANGYPKPLTKTOLSOGSPTPEAPGQGQVSSSLTKTKYTSALSODGVCEASNPL 528
OY	506	GKNLSLELELVNLTTLPOSNTTGTGSTASPHTRANSTPERKLREPESGVYIVAV 565
Db	539	G-----NIHHVFHEGTVAQT-----SQAGVAMVAV 554
OY	566	IYCIILVALIGAVLYFYKKGKLCRRSRGKOETLLPPSRKSELVEYKSCLKPEEMGLLO 625
Db	585	AISVALLLVLAIFYCMRRKRGPCQCQMG--EKSGSPPEGPLT--SHSGSQPREQGITLM 610
OY	626	GSS--GOKRAPDOGEK 640
Db	611	GSASGGAKHGSGGFGE 627
RESULT	13	
09JKB2		
ID	09JKB2	PRELIMINARY; PRT: 622 AA.
AC	09JKB2;	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)	
DT	01-OCT-2001 (TREMBLrel. 19, last annotation update)	
DE	LUTHERAN GLYCOPROTEIN.	
GN	LU.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	
CC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=129/SV;	
RA	Lee G., Willig T.-N., Parsons S.F., Anstee D.J., Mohandas N.,	
RT	Chasis J.A.;	
RL	"Mouse Lutheran Glycoprotein Gene."	
EMBL:	submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	
EMBL:	AF246667; AF61742.1; -.	

DR MGD: MGI:1929940; Lu.
DR Interpro: IPR003598; IG_C2.
DR Interpro: IPR003600; IG_Like.
DR Interpro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00406; IGC2; 2.
DR SMART: SM00410; IG_Like; 2.
DR Immunoglobulin domain.
KW SEQUENCE 622 AA; 67544 MW; 3A877738D96F66CD9 CRC64;

Query Match	18.4%	Score 618;	DB 11;	Length 622;
Best Local Similarity	28.5%;	Pred. No. 4e-42;		
Matches 191; Conservative	102;	Mismatches 275;	Indels 102;	Gaps 25;

QY	10	FLAACCCCPVAGVPGABDPAPAEIVLEVNGSTALLKGLCSLSOGSLHS--VDNF-----	63
Db	15	FLHS-----YSGAQLAHVSVPRREVMRGEOVALDC--TPREPREHYLDEMLVDG	65
QY	64	-----SVHKKEKTLTFRRVQGGSGSEPEGEQRLSLDRGRATIALTVPTQDRIPL	115
Db	66	TGARIRLASVEPQSEELGV-TSLGRVPRPEYDSR-----GRLLIAVQVGGDRDY	117
QY	116	CGG--RRRSQETRLQLVYKAPREPINQVNPGLGIPNNSKEPREVATCVARNQPIRQY	173
Db	118	CYVKKGAAGTSEANSSVVFAPREDTEVSPKGTLSVMDPFADELACSSNNNPVRILT	177
QY	174	WYKNRPLK---EKKRNHISQSTVESSGITLQSLQALVKEKDQAFUCELVYRP	230
Db	178	WYRNQGRLEVRMEVNGKQYITLPIVYREASGSLSTSLYLRLKDRDRADFHCANHYLP	237
QY	231	SSNNHK-ESREYTVDFVPTPKY--WLEVEYV--GMLKEQDREIKRLADGNPRPHFS	285
Db	238	SCQHRDLSSHPTFLYHPTHEVFERWGSPTTEGWRBEGDAVOLLCCGGSGSPSEYTF	297
QY	286	KQNPSTREAEETTDNNGVLEVPARKHSGRYEC--QGDLDTMISLSEPOELLVTV	343
Db	298	RQO-GTOE-EQLNVMLKGNLTLEGVNRQSGIYGCRAVEDYDADEVOYAK-LKLHAYL	354
QY	344	SVYKRSAPAPER---QGGSLTLTCAESQDDLEFQWLAEEFGQVLEBRPVQLDLKR	399
Db	355	DFLEIS-APELQVFLNSSLSTVNCSSRGLPTPTVWMTQDSV--TLADSPMLTDSVTF	410
QY	400	EAGGGYRCVASYPISPIGLNRLQLVNVAIFGPWM-----AFKERRKVVWKENVNLNSCEA	454
Db	411	DSAGITVCEASPTVPLLSPTQSQQLIYQAPRLKPRMEIMPKSGMSMTBEGDEVM-LTCSA	469
QY	455	SGHPRTTSMWNVNGTASQDDDPOR--VLSTLVNLYTPLELLEFGECTASNDLGKNTSL	512
Db	470	KFPFPKPLTWISQDGTAPAPPEFEGRGKMSSLAMKYVLSALSRVSCSEAINHGKKGNHF	529
QY	513	FLELVNLTLLPDSNTTGLSTSPASPHTRANSSTERKLLPEPESRCVVTVAVIYCLVY	572
Db	530	HGGSV-----APQTAQAGVANMAVAIVGILL	555
QY	573	AVLGAVLFLYKKGKLPQRSNGKOEITLPPRSKSELVEVYKSKLPRMGLOG--SSGD	630
Db	556	LLVVAFAFCMRKGRKPPGCCR--RAEKGAAPAREBEL--SHSGSERPENTGLMGSPGGG	611
QY	631	KRAPDQEK 640	
Db	612	KGGSGGFDE 621	
RESULT	14		
Q9R069			
ID	Q9R069	PRELIMINARY;	PRT; 622 AA.
AC	Q9R069;		
DT	01-MAY-2000 (Tremblrel. 13,	Created)	
DT	01-MAY-2000 (Tremblrel. 13,	Last sequence update)	
DT	01-DEC-2001 (Tremblrel. 19,	Last annotation update)	
DE	LUTHERAN GLYCOPROTEIN.		
GN	GPU OR LU.		

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rahuel C., Colin Y., Goossens D., Gane P., El Nemer W., Cartlon J.P.,
 RA Le Van Kim C.;
 RT "Characterization of a mouse laminin receptor gene homologous of the
 RT human blood group Lutheran gene";
 RL Immunogenetics 0:0-0(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Parsons S.F., Lee G., Chasis J.A., Tanner M.J.A., Anstee D.J.;
 RT "Mouse Lutheran glycoprotein";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21398054; Pubmed=11507772;
 RA Moulson C.L., Li C., Miner J.H.;
 RT "Localization of Lutheran, a novel laminin receptor, in normal,
 RT knockout, and transgenic mice suggests an interaction with laminin
 RT alphas in vivo";
 RL Dev. Dyn. 222:101-114(2001).
 DR EMBL: AF109160; AAF14226.1;
 DR EMBL: AF221507; AAF34657.1;
 DR EMBL: AF346663; AAK83237.1;
 DR MGI: 1929440; Lu.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig_5.
 DR SMART: SM00408; IGc2; 2.
 DR SMART: SM00410; IG_Like; 2.
 KW Immunoglobulin domain.
 SQ SEQUENCE 622 AA; 67670 MW; 257903F08D47EB4A CRC64;

Query Match 18.3%; Score 614; DB 11; Length 622;
 Best Local Similarity 28.4%; Pred. No. 8.5e-42;
 Matches 190; Conservative 102; Mismatches 276; Indels 102; Gaps 25;

QY 10 FILAACCCCPVAGPEAEAPAEVVEVGSALICKGLSGQNLSH--VDWF---- 63
 DB 15 FILSG-----YSGAQAELHVSVPPEVVRGEOVALDC--TPREHEHYVLEWFLVDG 65
 QY 64 -----SVHKEKRTLIFFRVGOGGSEPEYERLSLDRGATLALQVTPDERFL 115
 DB 66 TGAHRLASVEPQSEFLGTV-HSLGRVPEYVDSR-----GRLYIAKVQVGDGRDY 117
 QY 116 COGK--RPRSOERYIQLRVYKAPPEPNIOVPLGIPVNSKEPEVATCVGRNGYPIPOVI 173
 DB 118 CYYVKAAGAGTSATSSVAFATPEDETVSPNKGTLISVMDQAEIATCSSNNGNPVRIT 177
 QY 174 WYKNGRPLK---EKKNRVHIQSSQTVESGLTYLQSLIKAQVLEKDKDAQFYCELANRPL 230
 DB 178 WYKNGRQLLEVPMEVNOQGYITIRTVREASGLYSLTSLYLRKHDDRDANFHCAYHDLP 237
 QY 231 SGNHMK--ESRETVVVFYPTKEV--WLEVEPV--GMLEKGRVRIICLACGNPPPHSIS 285
 DB 238 SGGGRLDSTHFRILTHYPTHEHVEFWGSPSTEGWREGDAVQLQCGDGSFSPESFF 297
 QY 286 KQNSTREAEETNDNGVLVLEPARKHSGRYEC--QGIDLDTMISLSEPOELLVNY 343
 DB 298 RQO-GTQE-DQLNVLNKLGNLTIRVHNQGIYGCREDYDADAEVQLVYK-LKLHVAYL 354
 QY 344 SDVAVSPAAPER---QEGSSLITLCEASSQDLEFQMLKEETGOVLERGPVQLDLK 399
 DB 355 DPELIS--VEBELFVLINSSSTVAVNCARGLPTPTVAKWKDSV--TLAOGPMILSISVTF 410
 QY 400 EAGGRCVAVSVPISIPGLNLTQVLAIVAFEGPPM--AFKERRKYVKEENMLINISCEA 454
 DB 411 DSAGTVCYCASTPTVPLRLSTQSGQLVQGAPELKPNEIMPKGNSMTBEDVYM-LTCSA 469

QY 455 SGHPRPTISMNVMNGTASEODQDQOR--VLSTLNVLTPELLETGYECTASNDLCKNTSIL 512
 DB 470 RGFPEPKLTWSQGDPPAPPEPGEGRGKWSLSLVKVTALSREGVSCASNHGKRGV 529
 QY 513 FLELVNLTITLTPDSTNTTGLSTASTASPHTRANSTERTKLPPEESRGVYIVAVICVL 572
 DB 530 HFGSV-----APQTAQAVAAVAVSVGLL 555
 QY 573 AVIGAVLYFLYKKGKLPICRRSGKQETITLPPSRKSELVEVKSDKLPDEMGLQG--SSGD 630
 DB 556 LTVAAFYCMRRKRGRCR--RAEKGAAPPAREPEL--SHSGSERPEHTGILMGSPSGG 611
 QY 631 KRAPDQGEK 640
 DB 612 RGGSGGFDE 621

RESULT 15
 Q9ESS5 PRELIMINARY; PRT; 622 AA.
 AC Q9ESS5;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 GN LUTHERAN ANTIGEN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BA6/C;
 RA Taira E., Okumura S., Miki N.;
 RT "mouse Lutheran antigen";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB035511; BAB16053.1;
 DR InterPro: IPR003599; Ig_
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig_5.
 DR SMART: SM00409; IG_5.
 DR SMART: SM00408; IGc2; 3.
 DR SMART: SM00410; IG_Like; 3.
 KW Immunoglobulin domain.
 SQ SEQUENCE 622 AA; 67543 MW; 3264B7490E28428 CRC64;

Query Match 18.2%; Score 613; DB 11; Length 622;
 Best Local Similarity 28.4%; Pred. No. 1e-41;
 Matches 190; Conservative 102; Mismatches 276; Indels 102; Gaps 25;

QY 10 FILAACCCCPVAGPEAEAPAEVVEVGSALICKGLSGQNLSH--VDWF---- 63
 DB 15 FILSG-----YSGAQAELHVSVPPEVVRGEOVALDC--TPREHEHYVLEWFLVDG 65
 QY 64 -----SVHKEKRTLIFFRVGOGGSEPEYERLSLDRGATLALQVTPDERFL 115
 DB 66 TGAHRLASVEPQSEFLGTV-HSLGRVPEYVDSR-----GRLYIAKVQVGDGRDY 117
 QY 116 COGK--RPRSOERYIQLRVYKAPPEPNIOVPLGIPVNSKEPEVATCVGRNGYPIPOVI 173
 DB 118 CYYVKAAGAGTSATSSVAFATPEDETVSPNKGTLISVMDQAEIATCSSNNGNPVRIT 177
 QY 174 WYKNGRPLK---EKKNRVHIQSSQTVESGLTYLQSLIKAQVLEKDKDAQFYCELANRPL 230
 DB 178 WYKNGRQLLEVPMEVNOQGYITIRTVREASGLYSLTSLYLRKHDDRDANFHCAYHDLP 237
 QY 231 SGNHMK--ESRETVVVFYPTKEV--WLEVEPV--GMLEKGRVRIICLACGNPPPHSIS 285
 DB 238 SGGGRLDSTHFRILTHYPTHEHVEFWGSPSTEGWREGDAVQLQCGDGSFSPESFF 297

```

QY 286 KNPSTRAEETNDNGVILEPARKHSGRYEC--OGLDPMISLSPEOLLVNV 343
DB 298 RQO-CTQE-EQLNVNKGKGLTLEGVHRNOSGITGCRVEDYDADEVOQLVK-LKLVAYL 354
QY 344 SDVRVSPAPER---QEGSSLTITCEAESQDLEFQWLRETEGOVLERGPVQLHDLKR 399
DB 355 DPLEIS--VPELEFVFLNSSSTVNCARGLPTPARWTKDSV--TLADGPMLSLSQSVTF 410
QY 400 EAGGGYRCVAVSPSIPGLNRTOLVNAIFGPPM-----AFKERKVVYKENVVNLSCFA 454
DB 411 DSAGFYTCESASTPVPLLSRTQSFQLIVQGAPELKPNEIMPKSGNSWTEGDEVN-LTCSA 469
QY 455 SGHPRTISMNVNNGTASBDDQDOPOR--VLSTLNVLYTPELLETGVECTASNDLGKNTSIL 512
DB 470 RGFPEPKLWISORGDTPAPAPPEEGRGKSSSLMVKVTSALSREGVSCASNIHGKKGHYF 529
QY 513 FLELVNLTTLTPDSNTTGLSTASPHTRANSTSTERKLPPEPSRGVIVAVIYCIYL 572
DB 530 HRGSV-----APQTAQAGVAVMAVAVSVGLL 555
QY 573 AVLGAVLYFLYKKGKLPGRSGKQETLPPSRKSELVVEVKSDDLPEMGLLG--SSGD 630
DB 556 LLYVAAFYCMRRKGRPGCCR--RAEKAPPAPEL--SHGSEPREHTGLMGSPSGG 611
QY 631 KRAPDQGEK 640
DB 612 RGSSEGFGE 621

```

Search completed: June 28, 2002, 10:50:26
 Job time: 276 sec

Mon Jul 1 11:07:40 2002

us-09-653-961-2.rge

P:

TSITLPELVNTTITPDNSTTGTSTASPRANSTSTERKLPESRGVIVAVI
VCITIVAVLAVLVFLYKKGKLPKRSRSGQKITLPPSRKSELVVEKSKLPREMGLL
OGSSGPKRAPDGGGEXYIDLPH"

BASE COUNT 715 a 758 c 763 g 593 t
ORIGIN

alignment_scores:
Quality: 3359.00 Length: 646
Ratio: 5.200 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.845

alignment_block:
US-09-653-961-2 x BC006329

Align seg 1/1 to: BC006329 from: 1 to: 2829

1 MetGlyLeuProGlyLeuValCysAlaPheLeuLeuAlaLacCysCysC 17
17 ATGGGGCTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 66
17 scysProArGValAlaGlyValProGlyGluAlaGluInProAlaProG 34
67 CTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 116
34 IuLeuValGluValGlyValGlySerThrAlaLeuLeuLysCysGlyLeu 50
117 AGCTGTGAGGTGGAGAGTGGCGACGACGCGCTTCTGAACTGGGCTTC 166
51 SerGlnSerGlnGlyValAsnLeuSerHisValAspTyrPheSerValHisL 67
167 TCCAGTCCCAAGCACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 216
67 sGluLysArgThrLeuIlePheArgValArgGlnGlyGlnGlyInSerG 84
217 GGAGAGCGGAGCCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 266
84 IuProGlyGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 100
267 AACCTGGGAGTACGACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 316
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuLysG 117
317 CTGGCCCTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG 366
117 nGlyLysArgProArgSerGlnGlyValArgIleGlnLeuArgValTyrL 134
367 GGGCAAGCGGCTCGGTCCGAGGAGTACCGGATCCAGCTCCGCTGCTG 416
134 ysaLapProGluInProAsnIleGlnValAsnProLeuGlyIleProVal 150
417 AACCTCGGAGAGCCAAATCCAGGTCMAACCCCTGGGCACTCCCTG 466
151 AsnSerLysGluProGluGluValAlaThrCysValGlyValArgAsnGly 167
467 AACAGTAAGAGGCTGAGAGGTCGCTGCTGCTGCTGCTGCTGCTGCTG 516
167 rProIleProGluValIleTyrTyrLysAsnGlyArgProLeuLysGln 184
517 CCCCATCTCCAGTACTGCTGTAAGAGAGGCGGCTGCTGCTGCTGCTG 566
184 IuysAsnArgValHisIleGlnSerSerGlnThrValGlnSerSerGly 200
567 AGAAGAACCGGGTCCACATTCAGTCTCCAGACTGGAGAGCAGGCTG 616
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
617 TTGTACCTTCCAGAGATTTCTGAGGAGACGCTGTTTAAAGAGACAA 666
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyVal 234
667 AGATGCCAGTTTACTGTGAGCTCAACCTACCGGCTGCTGCTGCTGAG 716

234 IsMetLysGluSerArgGluValThrValProValPheTyrProThrGlu 250
717 ACATGAGAGAGTCCAGGAGAGTCCAGGCTGCTGCTGCTGCTGCTGCTG 766
251 LysValTyrPheGluValGluProValGlyMetLeuLysGluGlyAsp 267
767 AAAGTGTGGCTGGAAAGTGGACCCGTGGAAATGTGTAAGAGAGAGGAG 816
267 gValGluIleArgCysLeuAlaAspGlyValAsnProProPheAsnSer 284
817 CGTGGAAATCAGGTGTGGCTGCTGAGACCTGACCACTGACCACTGACCA 866
284 IeSerLysGlnAsnProSerThrArgGluAlaGluGluValThrThrAsn 300
867 TCAAGCAGACAGACCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 916
301 AspAsnGlyValAlaValLeuGluProAlaArgLysGluHisSerGly 317
917 GACAGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 966
317 gTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSerG 334
967 CTATGAAATGTGAGAGCTGAGCTTGGACACATATATATATATATATAT 1016
334 IuProGlnGluLeuLeuValAsnTyrValSerAspValArgValSerPro 350
1017 AACCAAGAGAACTACTGTGAACTATGTCTGACGCTCCAGAGTACGCTCC 1066
351 AlaAlaProGluArgGlnGluGlySerSerLeuThrThrCysGluAl 367
1067 GCAACCCCTGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1116
367 sGluSerSerGlnAspLeuGluPheGlnThrPheLeuArgGluThrGly 384
1117 AGAGAGTACAG 1166
384 IuysLeuGluArgIleProValIleGlnGluHisAspLeuArgGly 400
1167 AGCTGCTGGAAAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1216
401 AlaGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGlyLe 417
1217 GCAGAGAGGGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1266
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProThrMet 434
1267 GAAACCGACAGAGCTGATCAAGGAGGCAATTTTGGCCCTGCTGATGG 1316
434 IePheLysGluArgLysValTyrValLysGluAsnMetValLeuAsnLeu 450
1317 CATTCAG 1366
451 SerCysGluAlaSerGlyHisProArgProThrIleSerThrAsnVal 467
1367 TCTGTGAGAGGCTCAGGAGGACCCCGGCTGCTGCTGCTGCTGCTGCTG 1416
467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValIleuSerThr 484
1417 CGGACGCGCAGAGTAAACAGACCAAGATCCACAGAGAGAGAGAGAGAG 1466
484 uAsnValIleuValThrProGluLeuGlnGlnGlnGlnGlnGlnGln 500
1467 TGAATCTCTGCTGACCCCGGAGCTGTGGAGACAGGCTGTGAATACAG 1516
501 AlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluVal 517
1517 GCTCTCAAG 534
517 IAsnLeuThrThrThrThrProAspSerAspThrThrThrGlyLeuSer 534
1567 CAATTTACCAACCTTCAACAGAGCTCAACCAACCAACCAACCAACCAAG 1616
534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550

1617 CTTCCACCTCCGACCTCTCTATACAGACGACACCTCCACAGAGAGA 1666
551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIleVal 567
1667 AACCTCCGACGACGAGAGCCGGGCGCTGTCATCTGCTGTGATGCT 1716
567 LysIleLeuValIleValIleValIleValIleValIleValIleVal 584
1717 GTCCATCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1766
584 LysIleLeuValIleValIleValIleValIleValIleValIleVal 600
1767 AGAAGGAGCAAGCTGCGCTCAGCGCTCAGGAGAGAGATCAGCGCTG 1816
601 ProProSerArgLysSerGluLeuValValIleValIleValIleVal 617
1817 CCCCCTCTCTGTAAGAGCAACTGTTAGTAACTTAAGTAACTAAAGCT 1866
617 UPProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLysArgAlaP 634
1867 CCCAGAGAGATGGGCTCTCTGACGAGCAGCGGTGACAGAGAGGCTC 1916
634 roGlyAspGlnGlyLysTyrIleAspLeuArgHis 646
1917 CCGGAGACCAAGGAGAAATCATCATCTGAGGCGAT 1954
seq_name: gb_pr:AF089868
seq_documentation_block: 3335 bp mRNA linear PRI 04-MAR-1999
LOCUS AF089868 Homo sapiens cell surface glycoprotein P1H12 precursor, mRNA,
DEFINITION complete cds.
ACCESSION AF089868
VERSION AF089868.1 GI:4336423
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Gul, L., Chang, L., Browne, P.V. and Heibel, R.P.
TITLE P1H12 from human umbilical vein endothelial cells
JOURNAL Unpublished
2 (bases 1 to 3335)
REVIEW Gul, L., Chang, L., Browne, P.V. and Heibel, R.P.
AUTHORS Direct Submission
TITLE Submitted (02-SEP-1998) Medicine, University of Minnesota, Box 480
JOURNAL UMMC, 420 Delaware Street SE, Minneapolis, MN 55455, USA
FEATURES
source
1..3335
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="umbilical vein endothelial"
/note="Primary culture"
24..1964
/note="P1H12 antigen"
/codon_start=1
/product="cell surface glycoprotein P1H12 precursor"
/protein_id="PADI7799.1"
/db_xref="GI:4336424"
translation="MGILPRVCAFLAACCCPRVAVGPAEDAPABELVEVGSTAL
LILKGLSOSGSLSHVDFSVHREKRTLLIFRQOGGSEGEYORLSQDRATLA
LTVQTPQDERFLCOGRKPRSOEYRQLRVYAPPEPNIQVNPGLIVNSKEPEVAT
CVGRNGYPIPOVLYIKNGRPLKEKRVHIIOSOTVSSGLYLTKAKQIVVEDD
AOFYCELNRLRPSGNHMKESREVTVPFPTKVMLEVEPGMLKEGDRVEIRLAD
NPPEHFSIKONPSTRAEETNDNGVILVLPARKBSHGRVCOGLDOLVLEGPVL
POELLVNVSVSRVSPAPRROGSSLLTLCASOSOLEOMLREEDDOVLEGPVL
OLHLKREAGGTCVAVSPITGLRQIOLVAVLEGPMAAFKRYVMKENVNL
SCBASGHPRTISMNVGTSIPGDDOPQVLSLIVLVTPLELLEGVCAASNDLGN
TSLLPLVLTLPDSNTTGLSTVSPAPRANSTSRKLDPESRQVYVAVT
VCLIVLANICAVLTFLYKKKGLCRKSGKQETITLPPRSKSELVYVASKLIPKML
QGSQDKKAPEDQGEKITDLRH"

BASE COUNT 809 a 921 c 941 g 664 t
ORIGIN
alignment_scores:
Quality: 3356.00 Length: 646
Ratio: 5.203 Gaps: 0
Percent Similarity: 99.845 Percent Identity: 99.845
alignment_block:
US-09-653-961-2 x AF089868 ..
Align seg 1/1 to: AF089868 from: 1 to: 3335
1 MetGlyLeuProArgLeuValIleValIleValIleValIleValIleVal 17
24 ATGGGCTTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 73
17 scysProArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
74 CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
34 ILeuValGlyValGlyValGlySerThrAlaLeuLeuLysCysGlyLeu 50
124 ACCTGTGAGGTGAAGTGGGAGCAGACACCTTCTGAAAGTGGGCTC 173
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTyrPheSerValHisLy 67
174 TCCAGTCCCAAGCAACCTCAGCAGCTGACCTGCTTTCTGTCACAA 223
67 scLysArgThrLeuIlePheArgValArgGlnGlnGlnGlnSerG 84
224 GGAGAGAGGAGCGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 273
84 ILeuProGlyGluTyrGlnGluArgLeuSerLeuGlnAspArgGlyAlaThr 100
274 AACCTGGGAGTACGAGCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 323
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuGly 117
324 CTGGCCCTGACCTACAGTACCCCAAGAGCAGCAGCATCTCTGTGCTCA 373
117 nGlyLysArgProArgSerGlnGlyTyrArgIleGlnLeuArgValTyrL 134
374 GGGCAAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
134 yAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
424 AAGCTCCGAGAGGAGCAACATCCAGGTCACACCTGCTGCTGCTGCTGCT 473
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTyr 167
474 AACAGTAAAGAGCCTGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 523
167 rProIleProGluValIleTyrTyrLysAsnGlyArgProLeuLysGlu 184
524 CCCATCTCTCAAGTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573
184 IuLysAsnArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
574 AGAAGAACCGGCTCCACATTCAGTGTCCCAAGCTGGAAGTGAAGTGT 623
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValIleLysGluAsp 217
624 TTGTACACCTTCCAGAGATTCCTAAGAGCAGACCTGTTAAAGAGACAA 673
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsn 234
674 AGATGCCAGGTTTACTGCTGAGCTCAACTACCGGCTGCCAGGGAAC 723
234 IsetLysGluSerArgIleValThrValProValPheTyrProThrGlu 250
724 ACATGAGAGAGTCCAGGAGATCAACCTGCTGCTGCTGCTGCTGCTGCT 773

```

251 LysValItrPLeuGIuValGIuProValGIuMetLeuLYSGIuGIuAspAr 267
|||||
774 AAGGTGTGCTGGAAGTGAAGCCCGTGGGAATGCTGAAGGAAGGGAGCCG 823
|||||
267 GValGIuIleArGIuLeuAlaAspGIuAsnProProHIsPheSerI 284
|||||
824 CGTGGAAATCAGGTGTTGGCTGATGCAACCTCCACACACACTTCAGCA 873
|||||
284 ILeSerLYSGIuAsnProSerThrArgIuAlaGIuGIuGIuThrAsn 300
|||||
874 TCAGCAACGAGAAACCCAGCACAGGAGGACAGAGAGAGCAACAC 923
|||||
301 AspAsnGIuValLeuValLeuGIuProAlaArgLYSGIuHIsSerGIuAr 317
|||||
924 GACAGCGGGCTCTGCTGTGAGCCCTGCCGGAAGAAACAGTGGGGC 973
|||||
317 gTArgIuCYSGIuNGIuLeuAspLeuAspThrPheTLeSerLeuSerG 334
|||||
974 CTATGAATGTCAAGGCTGTGACTTGACACACATGATGCTGCTGAGTG 1023
|||||
334 IuProGIuGIuLeuLeuValAsnTrpValSerAspValArgValSerPro 350
|||||
1024 AACCAAGGAAGTACTGCTGTAATGTGTGACGCTCGAGTGAATGCC 1073
|||||
351 AlaAlaProGIuArgIuGIuGIuSerSerLeuThrLeuThrCYSGIuAl 367
|||||
1074 GCAGCCCTGAGAGACAGAGAGGACGACCTCACCTGACCTGTGAGGC 1123
|||||
367 AGIuSerSerGIuAspLeuGIuPheGIuTrpLeuArgIuGIuTrpArg 384
|||||
1124 AGAGATAGCCAGGACCTCGATCTCCAGTGTGAGAGAGAGAGACCC 1173
|||||
384 IuValLeuGIuArgIuProValLeuGIuHIsAspLeuLYSGIu 400
|||||
1174 AGTGTGTGAAGGGGCTGTGCTGCTGATGATGATGATGATGATGATG 1223
|||||
401 ALGGLGGLYLYTrArgCYValAlaSerValProSerIleProGIuLe 417
|||||
1224 GCAGAGGGGCTATCGCTGGTGGCTGTGTGCCAGATACCCGGCT 1273
|||||
417 uAsnArGIuTrGIuLeuValAsnValAlaIlePheGIuProProTrPhe 434
|||||
1274 GAACCCGACACAGCTGTGTCAACGTGGCATTTTGGCCCCCTGTGAGG 1323
|||||
434 ILePheLYSGIuArgLYSValItrPValLYSGIuAsnMetValLeuAsnLeu 450
|||||
1324 CATTCAGAGAGAGAGAGGTGTGGTGAAGAGATATGGTGTGAATCTG 1373
|||||
451 SerCYSGIuAlaSerGIuHIsProArgProThrIleSerTrpAsnVal 467
|||||
1374 TTTTGTGAAGCGTCAAGGGCACCCCGGCGCCACATCTCTGTGAACGT 1423
|||||
467 nGIuTrpAlaSerGIuGIuAspGIuAspProGIuArgValLeuSerTrp 484
|||||
1424 CGCAGAGCGAAGTGAACAAGACCAAGATCCACAGGATCTGTGAGCACCC 1473
|||||
484 euAsnValLeuValThrProGIuLeuLeuGIuThrGIuValGIuCYSThr 500
|||||
1474 TGAATCTCTGTGACCCCGGAGCTGTGAGACAGGTGTGATGACAGC 1523
|||||
501 AlaSerAsnAspLeuGIuLYSAsnTrpSerIleLeuPheLeuGIuLeuVal 517
|||||
1524 GCCCTCCAGACCTGGGCAAAAACACAGATCTCTCTCTGGAGCTGGT 1573
|||||
517 IAsnLeuThrTrpLeuThrProAspSerAsnTrpThrTrpArgLeuSer 534
|||||
1574 CAATTTAACCACTCTCACACAGACTCCAAACACACACTGGCTCAGCA 1623
|||||
534 hTrSerThrAlaSerProHIsThrArgValAsnSerThrSerThrGIuArg 550
|||||
1624 CTTCACCTCCAGTCTCTATACAGAGCCAAAGCACCTCCACAGAGAGA 1673
|||||

551 LysLeuProGIuProGIuSerArgIuValValIleValAlaValIleVal 567
|||||
1674 AAGCTGCCGAGCGGAGAGCCGGGCGCTGTCTCATCTGCTGTATATGT 1723
|||||
567 IValIleLeuValLeuAlaValLeuGIuValValLeuTrpPheLeuTrp 584
|||||
1724 GTGCATCTGTGCTCTGGCGGTGGCGGCGCTGTCTATTTCTCTATTA 1773
|||||
584 ySLYSGIuLYSLeuProCYSAArgArgSerGIuLYSGIuNGIuIleuHIs 600
|||||
1774 AGAAGGCGCAAGCTGCTGTGACAGCGCTCAGGAGAGAGAGATCACCGCT 1823
|||||
601 ProProSerArgLYSGIuLeuValValGIuValValSerAspLYSLys 617
|||||
1824 CCCCCTGTCTGTGAAGGCGAAGTGTATGTTGAAGTTAAGTCAATAGCT 1873
|||||
617 uProGIuGIuMetArgLYSLeuLeuGIuNGIuSerSerGIuAspLYSArgAla 634
|||||
1874 CCCAGAAAGATGGGCTCTCTGACAGGCGACAGCGGTGACAGAGGGCTC 1923
|||||
634 rGLYAspGIuNGIuGIuLYSTyrIleAspLeuArgHIs 646
|||||
1924 CGGAGACCGAGGAGAGAAATACATCATCTGAGGCAAT 1961
|||||

seq_name: gb_pat:AX331551
seq_documentation_block:
LOCUS AX331551 2943 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 2060 from Patent WO0194629.
ACCESSION AX331551
VERSION AX331551.1 GI:18122185
KEYWORDS
SOURCE
ORGANISM human.
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Euteleostomi; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Young,P.B., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Hortigan,S., Soppet,D.R. and Weiser,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A, 2060 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..2943
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 668 a 863 c 858 g 554 t
ORIGIN
alignment_scores:
Quality: 3327.00 Length: 646
Ratio: 5.166 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916
alignment_block:
US-09-653-961-2 x AX331551 ..
Align seg 1/1 to: AX331551 from: 1 to: 2943
1 MetGIuLeuProArgIuLeuValCYsAlaPheLeuLeuAlaCYsCY 17
|||||
8 ATGGGGCTTCCCGAGCTGTGCTGGCCCTTCTCTGCGCGCTGTGCTG 57
|||||
17 sCYsProArgValAlaGIuValProGIuLYSValAlaGIuGIuProAlaPro 34
|||||
58 CTGCTCGCGGTGGCGGTGGCGGAGAGGCTGAGAGAGCTGTGCGCTG 107
|||||
34 IuLeuValGIuValGIuValGIuSerThrAlaLeuLeuLYSGIuLeu 50
|||||
108 AGCTGTGAGGTGAGATGGGACAGACAGCCTTCTGAAGTGGCGCTC 157
|||||

```


51 SerGlnSerGlnGlnGlnLeuSerHisValAspTrpPheSerValHisIstly 67
158 TCACAGTCCCAAGGACCACTCAGCATGTCGCTGTTTCTGTCGCCAACAA 207
67 sglLysArgThrLeuIlePheArgValArgGlnGlnGlnIstGlnSerG 84
208 GGAGAAAGCGGACGCTCATCTTCCGTGTCGCGCAGAGGCGCAGGCGCAG 257
84 lUpProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
258 AACCTGGGGAGTACGACGACGAGGCTCAGCCTCCAGACGACGAGGCGCTACT 307
101 LeuAlaLeuThrGlnValThrProGlnAspGlnArgIlePheLeuGly 117
308 CTGGCCCTGACTCAAGTCACCCCGCCAGAGCGACATCTTCTGTGCCA 357
117 nglYlsArgProArgSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 134
358 GGGCAAGCGCCCTCGGTCCAGAGATCCGATCCAGCTCCGCGCTACAA 407
134 ysaIAProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 150
408 AAGCTCCGAGAGACCAACATCCAGGTCAACCCCTGGCATCCCTGTG 457
151 AsnSerLysGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 167
458 AACAGTAGAGAGCTGAGAGGTGCTGCTGCTGTAGAGAGAGAGGCGTA 507
167 rProIleProGlnValIleTrpTrpLysAsnGlnLysArgProLeuGlnGln 184
508 CCCCATCTCTCAAGTCATCTGTGACAAAGATGGCGGCTCTGAAAGAGG 557
184 lLysAsnArgValHisIleGlnSerSerGlnThrValGlnLysSerGly 200
558 AGAAGAACCGGGGTCCATTCATCTGCTCCCAACATCTGAGATCCAGTGGT 607
201 LeuTrpThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGlnAsp 217
608 TTGTACACCTTGACAGATATTCCAGAGCACAGCTGGTTAAAGAACAA 657
217 sAspAlaGlnPheTrpCysGlnLeuAspTrpArgLeuProSerGlnGln 234
658 AGATGCCAGTTTACTGTGAGCTCACTACCGGCTGCGCCAGCTGAGAAC 707
234 lMetLysGlnSerArgGlnValThrValProValPheTrpProThrGln 250
708 ACATGAGAGATCCAGGAGATGACCGCTCCCTGTTTCTTACCCACAGAA 757
251 LysValTrpLeuGlnValGlnProValGlnMetLeuLysGlnGlnAsp 267
758 AAGGTGTGGCTGAGAGTGAAGCCCGTGGAAATGCTGAAAGAGGAGCG 807
267 gValGlnLeuArgCysLeuAlaAspLysAsnProProHisPheSer 284
808 CGTGGAATCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 857
284 lSerLysGlnAsnProSerThrArgGlnAlaGlnGlnGlnGlnGlnGln 300
858 TCGCAGAGAGACCCAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 907
301 AspAsnGlnValLeuValLeuGlnProAlaArgGlnGlnGlnGlnGlnGln 317
908 GAAAGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
317 gTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 334
958 CTATGATATCTAGGCTGAGCTTGGACACATGATGCTGCTGCTGCTGCTG 1007
334 lUpProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 350
1008 AACCAAGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1057
351 AlaAlaProGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 367

1058 GCACGCCCTGAGAGACAGAGAGGCGAGCGCTCACCTGACCTGAGAGC 1107
367 agLysSerSerGlnAspLeuGlnPheGlnTrpLeuArgGlnGlnGlnGlnGln 384
1108 AGAAGTAGAGCAAGAGACCTGAGATCCAGTGGCTGAGAGAGAGAGAGAG 1157
384 lValLeuGlnGlnArgGlnProValLeuGlnLeuHisAspLeuLysArgGln 400
1158 AGGTGCTGGAAGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1207
401 AlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 417
1208 GCAGAGAGCGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1257
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlnGlnGlnGlnGlnGln 434
1258 GAACGCGACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1307
434 lAPheLysGlnArgLysValTrpValLysGlnAsnMetValLeuAsnLeu 450
1308 CATTCAG 1357
451 SerCysGlnLysSerGlnHisProArgProThrIleSerTrpAsnValAs 467
1358 TCTTGTAAAGCTCAGAGGACCCCGGCGCCACATCTCTGAGAGCTCAA 1407
467 nglYThrAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 484
1408 CGGACGCGCAAGTGAACAAAGACCAAGATCCACAGCGAGTCTGAGACACC 1457
484 euAsnValLeuValThrProGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 500
1458 TGAATGTCTCGTGAAGCCCGAGAGCTGTTGAGAGAGAGTGTAAATGCCAG 1507
501 AlaSerAspLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 517
1508 GCCTCCAGACAGCTGGCGCAAAACACAGATCTCTCTCTGAGAGCTGGT 1557
517 lAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlnLysSer 534
1558 CAATTTAACACCCCTCACACCGAGCTCCACACACACACACACACACACAC 1607
534 hSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGlnArg 550
1608 CTTCACGCTGCACTCTCTTACAGAGCCACAGACACCTCCACAGAGAG 1657
551 LysLeuProGlnProGlnSerArgLysValValIleValAlaValIleVal 567
1658 AAGCTGCGGAGCGGAGAGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCT 1707
567 lCysIleLeuValLeuValValLeuGlnValValLeuTrpPheLeuTrp 584
1708 GTGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1757
584 ysiLysGlnLysLeuProCysArgArgSerGlnGlnGlnGlnGlnGlnGln 600
1758 AGAAGGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1807
601 ProProSerArgLysSerGlnLeuValValValValValLysSerAspLys 617
1808 CCCCAGTCTGTAGAGCCAGACTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1857
617 uProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 634
1858 CCGCAGAAAGATGGGCTCTCTGAGGCGAGCGAGCGAGCGAGCGAGCGCTC 1907
634 roGlnAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 646
1908 CGGAGAGACAGGAGAGAAATACATGATCTGAGGCAAT 1945
seq_name: gb_pat:AX32735

Mon Jul 1 11:07:40 2002

us-09-653-961-2.rge

Page 6

```
seq_documentation_block:
LOCUS       AX332735                2943 bp    DNA    linear    PAT 09-JAN-2002
DEFINITION  Sequence 3244 from Patent WO0194629.
ACCESSION   AX332735
VERSION     AX332735.1
KEYWORDS    GI:18123369
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrihan,S., Soppet,D.R. and Weaver,Z.
TITLE       Cancer gene determination and therapeutic screening using signature
JOURNAL     Patent: WO 0194629-A 3244 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             1..2943
                    location/Qualifiers
BASE COUNT   668 a      863 c      858 g      554 t
ORIGIN
alignment_scores:
Quality: 3327.00      Length: 646
Ratio: 5.166          Gaps: 0
Percent Similarity: 99.690   Percent Identity: 98.916

alignment_block:
US-09-653-961-2 x AX332735
Align seg 1/1 to: AX332735 from: 1 to: 2943
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaAlaCysCysCY 17
|||||
8 ATGGGGCTCCAGAGCGGTGCTGCGCTTCTGCTGCGCGCTGCTGCTG 57
|||||
17 sCyProArgValAlaGlyValProGlyAlaGlyLeuInProAlaProG 34
|||||
58 CTGTCTCGCGTCCGCGGTGCTGCGCGAGAGCGTCAAGACCTGCGCTG 107
|||||
34 LuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
108 ACCTGCTGAGAGTGAGAGTGGCGACACAGCCCTCTGAAGTGGCGCTC 157
|||||
51 SerGlnSerGlnGlnAsnLeuSerHisValAspTyrPheSerValHisLy 67
|||||
158 TCCAGTGTCCAGAGCAACTCAGCCATGTGAGTGGTTTCTGTCACAA 207
|||||
67 sGlyLeuArgThrLeuIlePheArgValArgGlnGlyGlnGlnSerG 84
|||||
208 GGAGAGCGGAGCGCTCATCTTCCGTGTGCGCAGGCGCAGGCGCAGAGC 257
|||||
84 LuProGlyGlyTyrGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
258 AACCTGGGGAGTACGAGCAGCGCTCAAGCTCCAGAGACAGAGGGCTCT 307
|||||
101 LeuAlaLeuThrGlnValThrProGluAspGluArgIlePheLeuCysG 117
|||||
308 CTGGCGCTGACCTCAAGTCAACCCCGCAGAGAGCGCATCTTTTGTGCA 357
|||||
117 nGlyLeuArgProArgSerGlnGlyTyrArgGlnGlnLeuArgValTyrL 134
|||||
358 GGGCAGACGGCGCTCGTCCAGAGATACCGCATCCAGCTCCCGCTTACA 407
|||||
134 ySalAlaProGluLeuProAlaGlnGlnValAsnProLeuGlyLeuProAl 150
|||||
408 AAGCTCCGAGAGACCAACATCCAGTCAACCCCGGGGCAATCCCTGTG 457
|||||
151 AsnSerLysGluProGluLeuValAlaThrCysValGlyArgAsnGlyTy 167
|||||
```

```
458 AACAGTAAGACCTGAGAGAGTCCCTACCTGTGTAGAGAGAAACGGTA 507
167 rProIleProGlnValIleTyrTyrLysAsnGlyArgProLeuLysGluG 184
|||||
508 CCCATTCTCTCAAGTCACTGTGTACAAAGAAATGGCGGCTCTGAAAGAG 557
|||||
184 LuLysAsnArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
|||||
558 AGAAGAACCGGGTCCACTTCATGCTGCCAGATGTGGAGTGGAGTGTGT 607
|||||
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
|||||
608 TTGTACACCTTCCAGAGATTTCTGAAGGACAGCTGGTTAAAGAACACA 657
|||||
217 sAspAlaGlnPheTyrCysGluLeuAspTyrArgLeuProSerGlyAsn 234
|||||
658 AGATGCCAGTTTACTGTAGCTCAACTACCGGCTGCCAGTGGAGAAC 707
|||||
234 lsmLysGlnSerArgGluValThrValProValPheTyrProThrGlu 250
|||||
708 ACATGAAGAGATCCAGGAAATCACCGTCCGTTTCTACCCGACAGAA 757
|||||
251 LysValThrLeuGluValGluProValGlyMetLeuLysGluAspArg 267
|||||
758 AAGTGTGGCTGGAAAGTGAAGCCCGTGGGAATGCTGAAGGAAGGAGACG 807
|||||
267 gValGluIleArgCysLeuAlaAspGlyAsnProProProHisPheSer 284
|||||
808 CGTGGAAATCAGGTGTGTGGCTATGTGCAACCTCCACACTTTCAGCA 857
|||||
284 lSerLysGlnAsnProSerThrArgGluAlaGlyGluGluGluThrPro 300
|||||
858 TCAGCAAGCAGAACCCAGCAGCAGGAGAGAGAGAGAGAACAAACAAAC 907
|||||
301 AspAsnGlyValLeuValLeuGluProAlaArgTyrGlnLysSerGly 317
|||||
908 GACAAAGGGGCTCGTGTGTGGAGCCTGCCGGAAGAAACAGCTGGGG 957
|||||
317 gTyrGlnCysGlnLeuAsnLeuAspPheMetLysSerLeuLeuSerG 334
|||||
958 CTATGAATGTCCAGCGTGAAGTGGACACATGATTCCTGCTGATAG 1007
|||||
334 LuProGlnGluLeuLeuValAsnTyrValSerAspValArgValSerPro 350
|||||
1008 AACCAAGAGACTACTGTGTAATATGTCTGACGTCGACAGAGATCCC 1057
|||||
351 AlaAlaProGluArgGlnGlySerSerLeuThrLeuThrCysGluAl 367
|||||
1058 GCAGCCCTGAGAGACAGAGAGAGCAGACAGCTCAACCTGACCTGTAG 1107
|||||
367 agLysSerSerGlnAspLeuGluThrLeuThrLeuArgGluGluThrGly 384
|||||
1108 AGAGATAGCCAGACCTCGAGTCCAGTGGCTGAGAGAAAGAACAGAGC 1157
|||||
384 LuValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
|||||
1158 AGGTGTGAGAAAGGGGCTGTCTCTCACTGATCACTGAAAGAGAGAG 1207
|||||
401 AlaGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGly 417
|||||
1208 GCAGAGGGGGTGTACCTGTGGGTGTGGTGTGGCCAGCATCCCGGCT 1257
|||||
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTyrPhe 434
|||||
1258 GAACCGCACACAGCGTGTCAAGCTGGCATTTTGGGCGCCCTGTGATGG 1307
|||||
434 lAphLysGluArgLysValThrProValLysGluAsnMetValLeuAsnLeu 450
|||||
1308 CATTCAGAGAGAGAGAGTGTGGTGAAGAGAAATATGAGTGAATCTG 1357
|||||
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTyrAsnValAs 467
|||||
1358 TCTTGTGAAGCTCAGGGGACCCCGGCCACATCTCTGGAAGCTCAA 1407
|||||
```



```
67  sgluysarqthleuilephearqvalarglinsglinsglnserg 84
   |||
208  GGAGNAGCGACCTCATCTTCGTGGCCAGGCGGAGGCGAGAGGCG 257
   |||
84  lueprogluylutrglinsglinsglinsglinsglinsglins 100
   |||
258  AACCTGGGAGTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 307
   |||
101  leuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeu 117
   |||
308  CTGGCCCTGACTCAAGTACCCCAAGAGAGGAGGAGGAGGAGG 357
   |||
117  ngllysarqthleuilephearqvalarglinsglinsglins 134
   |||
358  GGGCAAGCGCCCTCGTCCAGAGTACCGCATCCAGCTCCGCTCA 407
   |||
134  ysaAlaProGlnGluProAsnIleGlnValAsnProLeuGlyIle 150
   |||
408  AACCTCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 457
   |||
151  AsnSerLysGluProGlnGluValAlaThrCysValGlyArgAsn 167
   |||
458  AACAGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 507
   |||
167  rProIleProGlnValIleTrpTyrLysAsnGlyArgProLeu 184
   |||
508  CCCCATTCTCTAAAGTACATCTGTGTAACAAGATGGCGGCTCT 557
   |||
184  luluysanargvalhisIleGlnSerSerGlnThrValGluSer 200
   |||
558  AGAAGAACCGGCTCCATTCAGTCCCTCCAGACGTGAGATCAG 607
   |||
201  leuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 217
   |||
608  TTGTACACCTTCAGAGTATTCGAAAGCAGCTGTTAAAGAGACA 657
   |||
217  saSpAlaGlnPheTyrCysGlnLeuAsnTyrArgLeuProSer 234
   |||
658  AGATGCCAGTTTACTGTGAGCTCAACTACCGGCTGCCAGTGG 707
   |||
234  lsmetLysGluSerArgIleValThrValProValPheTyrPro 250
   |||
708  ACATGAGAGGAGTCCAGGAGTCAAGCTCCGTCTTCTACCGACA 757
   |||
251  lyeValTrpLeuGlnValGluProValGlyMetLeuLysGln 267
   |||
758  AAGTGTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 807
   |||
267  gValGluIleArgCysLeuAlaAspLysAsnProProIlePhe 284
   |||
808  CGTGGAAATCAAGCTTTGGCTGATGGCAACCTCCACACACTCA 857
   |||
284  lseSerLysGlnAspProSerThrArgIleAlaGlnGluLys 300
   |||
858  TCAGCAAGCAGAACCCAGCAAGGAGGAGGAGGAGGAGGAGGAG 907
   |||
301  AspAsnGlyValLeuValLeuGluProAlaArgLysGlnHis 317
   |||
908  GACAGAGGAGTCTGCTGCTGAGACCTCCGAGAGAGACACATG 957
   |||
317  gTyrGlnCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 334
   |||
958  CTAATGATGTCAGGCTGAGACTTGAGACCATGATATCGGTGAG 1007
   |||
334  lueProGlnGluLeuLeuValAsnTyrValSerAspValArg 350
   |||
1008  AACCAAGAGAACTACTGCTGACTATGCTGACGTCCGAGTACGCC 1057
   |||
351  AlaAlaProGlnLutArgGlnGlnLysSerSerLeuThrLeu 367
   |||
1058  GCAGCCCTGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAG 1107
   |||

367  aGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGlnGlu 384
   |||
1108  AGAGATAGCCAGAGACCTGAGTCCAGTGGCTGAGAGAGAGAG 1157
   |||
384  lueValLeuGlnArgGlyProValLeuGlnLeuHisAspLeu 400
   |||
1158  AGGTCTGAGAAAGGGGAGCTGTGCTTCAAGTTCAGTCACTGA 1207
   |||
401  AlaGlyGlyGlyTyrArgCysValAlaSerValProSerIle 417
   |||
1208  GCAGAGGAGGAGTACGCTGCTGCTGCTGCTGCTGCTGCTG 1257
   |||
417  uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProPro 434
   |||
1258  GAACCGCACACAGCTGCTCAAGCTGGCCATTTTGGCCCTTGA 1307
   |||
434  lA PheLysGluArgLysValTrpValLysGluAsnMetVal 450
   |||
1308  CATTCAGAGAGAGAGAGTGTGTGTAAGAAGATATGCTGTAAT 1357
   |||
451  SerCysGluAlaSerGlyHisProArgProThrIleSerTrp 467
   |||
1358  TCTTGTGAAGCGTCAAGGAGCACCAGGACCATCTCTCGAAG 1407
   |||
467  ngllyThrAlaSerGlnGlnAspGlnAspProGlnArgVal 484
   |||
1408  CGGACGGGCAAGTGAACAGAGCAAGATCCACAGCGAGTCTG 1457
   |||
484  euAsnValLeuValThrProGlnLeuLeuGlnTrpGlyVal 500
   |||
1458  TGAATGCTCTGCTACCCCGAGCTGTGGAGAGAGGAGTGTGA 1507
   |||
501  AlaSerAsnAspLeuGlnLysAsnThrSerIleLeuPheLeu 517
   |||
1508  GCCTCAACAGCAGCTGGGCAAAACACAGCATCTCTCTGAGAG 1557
   |||
517  lAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyLe 534
   |||
1558  CAATTATACACCCCTCACACAGACTCCACACACACAGCTGCT 1607
   |||
534  hrSerThrAlaSerProHisThrArgAlaAsnSerThrSer 550
   |||
1608  CTTCCACGCCAGTCCCTATACAGAGGAGGAGGAGGAGGAGG 1657
   |||
551  lysLeuProGluProGlnSerArgLysValValIleValAla 567
   |||
1658  AAGCTCCGAGAGCGGAGAGCCGGGCTGTGCTATGCTGCTAT 1707
   |||
1708  GTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1757
   |||
584  yAlaGlyLysLeuProCysArgArgSerGlyLysGlnGlu 600
   |||
1758  AGAAGGAGAGCTCCGCTGAGGAGGAGGAGGAGGAGGAGGAG 1807
   |||
601  ProProSerArgLysSerGlnLeuValValGluValLysSer 617
   |||
1808  CCCCGCTCTGTAAGACCGAAGTGTATGTAAGTCAAGATTA 1857
   |||
617  uProGlnGlnMetGlyLeuLeuGlnGlySerSerGlyAspLys 634
   |||
1858  CCCAGAAAGATGAGGCTCTCTGAGGAGGAGGAGGAGGAGG 1907
   |||
634  roGlyAspGlnGlyGluLysTyrIleAspLeuArgHis 646
   |||
1908  CCGGAGACCAAGGAGAAATACATCATCTGAGGAGT 1945
   |||

seq_name: gb_pr:U00MUC18B
seq_documentation_block:
LOCUS      HUMMUC18B              3306 bp    mRNA    linear
DEFINITION Human MUC18 glycoprotein mRNA, complete cds.
```

ACCESSION M28882
VERSION M28882.1 GI:529723
KEYWORDS MUC18 glycoprotein; cell adhesion molecule; immunoglobulin-like
sequence; integral membrane glycoprotein.
SOURCE Human cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3306)
AUTHORS Lehmann, J.M., Riethmuller, G. and Johnson, J.P.
TITLE MUC18, a marker of tumor progression in human melanoma, shows
sequence similarity to the neural cell adhesion molecules of the
immunoglobulin superfamily
Proc. Natl. Acad. Sci. U.S.A. 86 (24), 9891-9895 (1989)
JOURNAL MEDLINE 90093968
REFERENCE 2 (bases 1757 to 1945)
AUTHORS Sers, C., Kirsch, K., Rothbacher, U., Riethmuller, G. and Johnson, J.P.
TITLE Genomic organization of the melanoma-associated glycoprotein MUC18:
implications for the evolution of the immunoglobulin domains
Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)
JOURNAL MEDLINE 93391384
REFERENCE 3 (bases 1 to 3306)
AUTHORS Johnson, J.P.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1989) Judith P. Johnson, Institute for
Immunology, University of Munich, Goethestrasse 31, Munich, Germany
80336
COMMENT On Aug 15, 1994 this sequence version replaced gi:188861.
FEATURES
source
location/Qualifiers
1..3306
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="zapyl, drop 1, drop 4, drop 4.2, drop 4.7, drop
4.11"
/cell_line="Mel Juso"
/tissue_type="melanoma"
8..91
8..1948
/codon_start=1
/product="MUC18 glycoprotein"
/protein_id="AAA20922.1"
/db_xref="GI:529724"
/translation="MGLPRLVCAFLAACCCPRVAGVPEADQPAELVEVEVGSTA
LAKGSLGSGNSLHVDFVHKERKTLIRVVGQSGSPGEVEQSLSDGATLA
LTVYTPQDERFLFCGRPSQERQIRLVKAPEDENIQVNPGLIVNKEEVEAT
CVGNRGPIQVITWKYKGRPLKEKNVHTIOSQTVSGSLYTLQSLKQVKEKD
AQTCELNIRLPSGNHMKESRETVVPTFYTEKVMLEVEVGHAKESDREYICLADG
NPPEFISIKQNPSTRAEETINDNGVLVEPARKHSRYECQANNDVLTSLISE
POELLVYSDVRYVSPAAPERQSGSLTICESSDLEFQWLRETTDVLERGPLY
QLHLKREAGGGRVCAVSIPGLNRTQVLAIFGPMMAKERKVMKENVLNL
SCESGHPRTISWNVGTASEDQDQRLVLAIVLPPELLETGEVCAASNDLGN
TSILFLVNTLTTPDSNNTTGSTASTASPHTPANSTERKLPEESGVYIVAVI
VCTLVAVLGVLEFYKKKRLPCRRSGKQETILPSPRKELVVEVKSDDLPEMGL
QSGSGDKRAPEGDEKTIIDLRH"
mat_peptide
92..1945
/product="MUC18 glycoprotein"
misc_feature
122..373
/note="immunoglobulin-like (V set) domain I"
misc_feature
467..694
/note="immunoglobulin-like (V set) domain II"
misc_feature
795..988
/note="immunoglobulin-like (C2 set) domain"
misc_feature
1076..1249
/note="immunoglobulin-like (C2 set) domain"
misc_feature
1337..1525
/note="immunoglobulin-like (C2 set) domain"
misc_difference
1766
/note="g is deleted"
misc_difference
1880-1881
/replace="g"
misc_difference
/citation="insert g"
misc_difference
/citation="[1]"

polYA_signal 3273..3278
BASE COUNT 799 a 916 c 924 g 667 t
ORIGIN
alignment_scores:
Quality: 3327.00 Length: 646
Ratio: 5.166 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916
alignment_block:
US-09-653-961-2 x HUMMUC18B ..
Align seg 1/1 to: HUMMUC18B from: 1 to: 3306
1 MEGGLYLEUPROARGLEUVALCYSAIAPHLEULEUVALALACYSYSCY 17
|||||
8 ATGGGGCTTCCAGCGCTGGCTGCGCTTCTGCGCGCTGCTGCTG 57
|||||
17 SCYSPARGVALALAGLYVALPROGLYGLUALAGLUGINPROVALPROG 34
|||||
58 CTGTCTCCGCGTCCGGGTGTCGCCGAGAGCTGAGCAGCTCGCCCTG 107
|||||
34 IULEUVALGLUVALGLUVALGLYSERTHALALEULEULYSYSGLYLEU 50
|||||
108 AGCTGTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 157
|||||
51 SERLINSENGINGLYASNULEUSERHISVALASPTIPHESEVALHISLY 67
|||||
158 TCCAGTCCCAAGGACCACTGACGATGTGACTGTCTTCTGTCCACAA 207
|||||
67 SGLUYSARGTHREULEULEPHEARGVALARGINGLYINGLYINSENG 84
|||||
208 GGAGAGCGGAGCGCTCATCTTCCGTGTGCGCCAGGCGCAGGCGCAGGCG 257
|||||
84 IUPROGLYGLUTYRGLUGLARGLEUSERLEUGLNASPARGGYALATHR 100
|||||
258 AACCTGGGAGTACGACGACGCGCTCAGCTCCAGGACAGAGGGGCTACT 307
|||||
101 LEUVALLEUTHRGINVALTHPRGILNASPGLUARGILEPHELEUCYSGI 117
|||||
308 CTGGCCCTGACTCAAGTCACCCCAAGACGAGCGCATCTTCTGTGCCA 357
|||||
117 NGLYLSARGPROARGSERGLINGLYTYRARGILEGLLEUARGVALTYRL 134
|||||
358 GGGCAAGGCGCCCTGCTCCAGAGTACCGCATCCAGCTCGCGCTACAA 407
|||||
134 YSALAPROGLUGLUPROASNILEGLINVALASNPROLEUGLYLEPROVAL 150
|||||
408 AAGCTCCGAGGAGCCAAACATCCAGGTCAACCCCTGGGCAATCCCTGTG 457
|||||
151 ASNSERLYSGIUPROGLUGLVALALATHCYSAIAGLYARGASNGLYTY 167
|||||
458 AACACTAGGAGCCCGAGAGGTGCTACTACGTGTAGGAGGAAGAGGCTA 507
|||||
167 TPRIILEPROGLINVALIETTPYLYLSASNGLYARGPROLEULYSGLUG 184
|||||
508 CCCCATTTCTCAAGCATCTGTGTACAAAGAGGCGGCTGTGAAGAGG 557
|||||
184 IULYSASNARGVALHISILEGINSESERGINTRVALIGLUSERSEGLY 200
|||||
558 AGAAGAACCGGGTCCACATTTCAGTCGTCCGAGCTGAGTGCAGTGCAGT 607
|||||
201 LEUTYRTHREUGLINSERILEULEYSAIAGINLEUVALLYSGIUAAPLY 217
|||||
608 TGTACACCTTGACAGATATTCTGAAGCAGACAGCTGTGTAAGAGCAA 657
|||||
217 SASPALAGINPHERYCYSGIULEUASNTYRARGLEUPROSERGLYSNH 234
|||||
658 ACATGCCAGTTTACTGTGAGCTCACTACCGGCTGCGCAGTGGGAACC 707
|||||
234 ISMELLYGLUSERARGSLUVALTHRVALPROVALPHETYRPROTHRGLU 250
|||||

```

708 ACATGAAGAGTCCAGGAAGTCACCGCTCTTTTCTACCCGACAGAA 757
251 LysValTrpLeuGluValGluProValGlyMetLeuLysGluGlyAsp 267
758 AAAGTGTGGCTGGAACTGGAGCCCGTGGAAATGCTGAAGAAAGGGGACCG 807
267 gValGluLeuArgCysLeuAlaAspGlyAsnProProHisPheSerI 284
808 CGTGAATATCAGGTGTGTGGCTGATGCAAAACCTCCACACACTTCAGCA 857
284 LeSerLysLysAsnProSerThrArgGluAlaGluGluThrThrAsn 300
858 TGAAGAGAGAAACCCACACAGGAGGAGAGAAAGAACAAACCAAC 907
301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGly 317
908 GACAAAGGGGCTCTGGTGTCTGGAGCTCCCGGAAGAAACAAAGTGGCG 957
317 gTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSer 334
958 CTATGATGTACAGCCCTGGAACTGGACACCATGATCGCTGTGAGTG 1007
334 LProGlnGluLeuLeuValLeuValLeuValLeuValLeuValLeu 350
1008 AACCAAGAGAACTACTGTGAACTATGTCTGTGACGTCCGAGTGAAGTCC 1057
351 AlaAlaProGluLysGlnGluLysSerSerLeuThrLeuThrCysGlu 367
1058 GCAGCCCTCGAGAGACAGGAAGGACGACGCTCACCCTGACCTGTGAGGC 1107
367 LeuLeuSerSerGlnAspLeuGluPheGluTrpLeuArgGluThrGly 384
1108 ACAGAGTACCCAGGACCTGAGTGTCCAGTGTGCTGAGAGAGAGACAGCC 1157
384 LValLeuGluLysArgGlyProValLeuGlnLeuHisAspLeuLysArg 400
1158 AGCTCTCGAAAGGGGCGCTGTCTGCTTCACTGATGACCTGAAGGGGAG 1207
401 AlaGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGly 417
1208 GCAGAGAGCGGCTATCGCTGCTGCTGTGCTGACCATACCCGGGCT 1257
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTrpMet 434
1258 GAACCGGACACAGCTGTGTCAAGCTGGCCATTGGTGGCCCTTGGAGAG 1307
434 LAlaPheLysGluArgLysValTrpValLysGluAsnMetValLeuAsn 450
1308 CATCAAGAGAGAGAAAGTGTGGTGAAGAAAGAAATATGTGTGATCTG 1357
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTrpAsnVal 467
1358 TCTTGTGAAGGTCTCAGGGCACCCCGGCCACCATCTCTCGAAGCTCA 1407
467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThr 484
1408 CGGCGCGCAAGTGAACAAAGACCAAGATCCACAGGAGTCTTGAGACCC 1457
484 euAsnValLeuValThrProGluLeuLeuGluThrGlyValGluCysThr 500
1458 TGAATGTCTCTGTGACCCCGAGCTGTGGAGACAGGTGTGATGACAG 1507
501 AlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluLeu 517
1508 GCCCTCAACAGCCTGGGCAAAAACACAGCATCTCTTCCGAGAGCTGT 1557
517 LAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSer 534
1558 CAATTAAACCCCTCACACACAGACTCCAAACACACACATGGGCTCAGCA 1607
534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
1608 CTTCCACAGTCCAGTCTCATACAGAGCAACACACACTCTCCACAGAGAGA 1657

```

```

551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIle 567
1658 AAGCTCCGAGCCGGAGAGCCGGGGCGGTGTCTATCGTGTGATGTAT 1707
567 LysIleLeuValLeuAlaValLeuGlyAlaValLeuThrPheLeuThr 584
1708 GTGCATCTGTCTCTGGCGGTGTCTGGGCGCTCTTATTTCTCTATA 1757
584 yslYsglyLysLeuProCysArgArgSerGlyLysGlnIleThrLeu 600
1758 AGAAGGGCAAGCTCCGTGCAGCGCTCAGGAGAGAGAGATCAGCCTG 1807
601 ProProSerArgLysSerGluLeuValValGluValLysSerAspLys 617
1808 CCCCCTCTCTTAAGACCGAATCTGTAGTTGAAGTAAATCAATTAAGCT 1857
617 uProGluGluMetGlyLeuLeuGlnGlnGlySerSerGlyAspLysArg 634
1858 CCCAGAAAGATGGGCTCTCGCAGGGGACACAGCGGTGACAAAGGCGCT 1907
634 roGlyAspGlnGlyGluLysTyrIleAspLeuArgHis 646
1908 CGGAGAGACAGGGAGAGAAATACATCATCTGAGGAT 1945

```

seq_name: gb_ro:AB035508

seq_documentation_block:

LOCUS AB035508 2890 bp mRNA linear ROD 03-OCT-2000

DEFINITION Mus musculus mRNA for 1-glicerol/MUC18, complete cds.

ACCESSION AB035508

VERSION AB035508.1 GI:10566952

KEYWORDS 1-glycerin/MUC18.

SOURCE Mus musculus (strain: BALB/c) male cDNA to mRNA.

ORGANISM Mus musculus (strain: BALB/c) male cDNA to mRNA.

REFERENCE 1 (bases 1 to 2890)

AUTHORS Taira, E., Okumura, S. and Miki, N.

JOURNAL Published Only in Database (2000) In press

REFERENCE 2 (bases 1 to 2890)

AUTHORS Taira, E., Okumura, S. and Miki, N.

JOURNAL Direct Submission

REFERENCE 3 (bases 1 to 2890)

AUTHORS Taira, E., Okumura, S. and Miki, N.

JOURNAL Submitted (03-DEC-1999) Eiichi Taira, Osaka University Medical School, Department of Pharmacology, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan (E-mail: etaira@pharm.med.osaka-u.ac.jp, Tel.: +81-6-6879-3521, Fax: +81-6-6879-3521)

FEATURES

source

1..2890

/organism="Mus musculus"

/strain="BALB/c"

/db_xref="taxon:10090"

/sex="male"

20..1966

/gene="1-glicerol/MUC18"

/product="1-glicerol/MUC18"

/protein_id="BAB16050.1"

/db_xref="GI:10566953"

/translation="MGLPKLYCVLFPAACCCRRRAAGVPEEKOPVTPDIVEAEGS

TALCKGPRBSAGNFSQVDMFLIKERQILIFRHHQGGREGEYRRLSDSVAT

LALHVPIDHRMFLCKSKPRLODHYELVPEKAVVEPVGLKEGDVIRILT

ATCYGRNGYPTPOLYWKNSPLDEENRHHIOSQIVSSGLTLKSVLSARLVED

KDAQFCELSLRPSGNHMKESKEVTVPEVPAKVVVEPVGLKEGDVIRILT

DGNQPFHTINKKDPSTGEEMESTDENGILSEPAKHHSGLYOCOSIDLETITLS

SDPLELVNVYSDVQVNPATPEVOGESLITLCEASNDLREEMLRDKTGVQENAVP

VLOLNVRREAGGRYLCMASVPRVGLNRTOLYSGIFGSPMALEKRWKQVQENAVP

NLSCGASHQPTISMVNGSATENMPDPTQVSTLVYTPPLLETGACACTSNLSG

SNITTYLKLVTITLTLLPDSQTTGLSTLTLSHTRANSTSTKRLPQPSKGVIVA

VIVCTILAVLGAALVFLYRKGLKPCGRSGKQBITLTPRKSEFVEVSKDLPEEMA

BASE COUNT 708 a 735 c 783 g 664 t
ORIGIN
LLOGSNDKRAPGDQSEKYLDRH"
alignment_scores: Quality: 2642.00 Length: 648
Ratio: 4.374 Gaps: 1
Percent Similarity: 93.210 Percent Identity: 76.389
alignment_block:
US-09-653-961-2 x AB035508 ..
Align seg 1/1 to: AB035508 From: 1 to: 2890
1 MetGlyLeuProArgLeuValAlcysAlaPheLeuLeuAlaAlaCysCysCyl 17
|||||
20 ATGGGGCTGCCCAACTGGTGGCTCTTGTGCTGCTGCTGCTGCTGCTGCTG 69
17 scysProArgValAlaGlyValProGlyGluAlaGluGln.....ProA 32
|||||
70 CTGTGCCCTGCCCGGGGTGCTGCCAGAGAGAGAAAGCAGCATACCA 119
32 laProGluLeuValGluValGluValGlySerThrAlaLeuLeuLeuScys 48
:|||||
120 CCCCCCAGCTGGTGGAGCAGAAAGTGGGCGACAGACCCCTCTCAAGTGT 169
49 GlyLeuSerGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 65
|||
170 GGGCCCTCAGGGGCTCAGGAGCACTCAGCACTCAGCACTCAGCACTGAT 219
65 HisGlyGluGlyArgGlyLeuLeuPheArgValArgGlnGlnGlnGlnG 82
:|||||
220 TCACAAGAGAGAGCAATGATGATTTCCGTGTGACACCAAGCAGGAGGCC 269
82 InsGlnProGlyGlyGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGly 98
|||
270 AAGCGGAACTGTGTAATATGAGACCGCGCTTAAAGCTTCAAGACACGCTG 319
99 AlaThrLeuAlaLeuThrGlnValThrProGlnAspGluArgGlyLeuPhe 115
|||||
320 GCTACTGTGGCCCTGAGTCACGTCTCCCTCCATGATGAGCAATTTCTCT 369
115 uCysGlnGlyLysArgProArgSerGlnGluGlyArgGlnGlnGlnGlnG 132
|||
370 GTTAAAGAGCAAGCAGCAGCGCTCAGAGATCAGTACAGTTCAGCTTCA 419
132 alTyrLeuAlaProGlnGluProAsnGlnGlnValAsnProLeuGlyLe 148
|||||
420 TCTTCAAGACCCAGAGAGCACTATTCAGCAATGTCGTGGGCATTC 469
149 ProValAsnSerLysGluProGlnGluValAlaThrCysValGlyArgAs 165
|||||
470 CATGTGACAGCAGCAGAGCTCAGGAGGTGCTGCTGCTGCTGCTGCTG 519
165 nGlyLysProLeuProGlnValLeuTyrLysAsnGlyArgProLeuL 182
|||||
520 CCGCTACCCCAATCTCAAGTCCATGTGTACAGAAACAGTCCGCTTGC 569
182 ysgLgLuLysAsnArgValHisIleGlnSerGlnThrValGlnSer 198
:|||||
570 AAGAGAGAGAGAGAGAGTTCATATCCAGTATCAGCATTTGTGAGTCC 619
199 SerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
|||||
620 AGTGGCTGTACACCTGAAGAGTCTCTGAGTGACAGCCTAGTTAAGA 669
215 uAspLysAspAlaGlnPheLysGlyGluLeuAsnTyrArgLeuProSerG 232
|||||
670 AGACAAAGATGCCAGTTTACTGTGAACTCAGTACCGGCTTACCCAGTG 719
232 lAsnHisMetLysGluSerArgGluValThrValProValPheTyrPro 248
|||||

720 GGAACACATGAGAGAACTTAAGAGGCTCACTGCTCTTTTCTTACCT 769
249 ThrGluLysValITrPLeuGluValGluProValGlyMetLeuLysGluG 265
:|||||
770 GCAGAAAAAGTGGGTGAGAGCTAGAGCTGTGGGGCTGCTGAAGAGAG 819
265 yAspArgValGluLeuArgCysLeuAlaAspGlyLysAsnProProHisP 282
|||||
820 GGATCATGTGACAAATCAGTGTGTGACAGATGGCAGACCCCTCAACCC 869
282 heSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluGluThr 298
|||||
870 TCACATATCAACAGAGAGAGACCCAGCAGCTGGGGAGATGAGAGAGAG 919
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGlnHis 315
:|||||
920 ACCGATGAATAATGGGCTCCTGCTGAGAGCTCCGAAAAAGCAGCATAG 969
315 rGlyArgTyrGlyGlyGlnGlyLeuAspLeuAspThrMetIleSerLeu 332
|||||
970 CGGGCTTACCAAGTGTGAGAGTGTGAGCTGGAAGAACTACCATCAGCT 1019
332 euSerGluProGlnGluLeuLeuValAsnTyrValSerAspValArgVal 348
|||||
1020 CAATGTACCCCTGAGCTGTGTGTGATGATGATGATGATGATGATGATG 1069
349 SerProAlaAlaProGluArgGlnGlySerSerLeuThrLeuThrCys 365
:|||||
1070 AATCAACTGCCCCCTGAAGTCCAGAGAGGTGAGAGCTCAGCTGAGCTG 1119
365 sGluAlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGluGlu 382
|||||
1120 CGAGAGCAAAAGTACACAGCAGCTTGTGATGATGATGATGATGATGAT 1169
382 hrGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLys 398
:|||||
1170 CAGCCAGCTGCTGGAGAAAGGTCCCTCCTCAGTAAACACAGTGAAG 1219
399 ArgGluAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIleP 415
|||||
1220 CGGAAACAGAGGGAGCGTATCTGTGATGCAATGCTCCCGAGAGTTC 1269
415 oGlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProPro 432
|||||
1270 TGGCTTAATCGTACACCGAGCTGTGAGCGGTGATTTTGGGTCCCAAT 1319
432 rPheAlaPheLysGluArgLysValITrValLysGluAsnMetValLeu 448
|||||
1320 GGATGGCATTAAGAGAGAGAGGTGTGGGTCCAGAAATGCAGTGTG 1369
449 AsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSerTrpAs 465
|||||
1370 AATGTGCTTGTGGGCTTCAAGACATCTCAGCCACCACTTCCCTGGAA 1419
465 nValAsnGlyThrAlaSerGlnGlnAspGlnAspProGlnArgValLeu 482
|||||
1420 TGTCAATAGTTTGGCAACGTAAATGAACCAATCCAGAGACAGTGTGTA 1469
482 erThrLeuAsnValLeuValIThrProGluLeuLeuGluThrGlyValGlu 498
|||||
1470 GCACCTTGAATGTCTTGTGACCCAGAGCTTCTGAGACAGTGTGACAG 1519
499 CysThrAlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeu 515
|||||
1520 TGTACAGCCCTCCAACTCCCTGGGCTCAACACACCAACCATTTGTTGAA 1569
515 uLeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyL 532
|||||
1570 GCTGTACACTTAAACACCTCATACCTGATCAGCAGCAACACCATGGCC 1619
532 euSerThrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThr 548
|||||
1620 TCAGCACCTCTACAGTCACTGTACACCAAGAGCCAAAGCAGCTTCACA 1669

```
549 GluArgLysLeuProGluProGluSerArgGlyValValIleValAlaIaIa 565
|||||
1670 GAGAAAGAGCTGCCACAGCAGAGCAAGAGTGTGTCATCTGCTGCTGCT 1719
565 ILeValGlyIleLeuValIleAlaIaValGlyValAlaValLeuTyrPheI 582
|||||
1720 GATAGTGTGTACCTGTGTGCTGTGCTGTGCTGCTGCTCTCTATTTTCC 1769
582 eUtyrLysGlyLysLeuProCysArgArgSerGlyLysGlnGluIle 598
|||||
1770 TCTACAAGAGGCGCAAGCTGCCATGTGACGCTCGGAAACAGAGATC 1819
599 ThrLeuProProSerArgLysSerGluLeuValValGluValLysSerAs 615
|||||
1820 ACGCTGCCCCCGCACTCGTAAGAGTGAATTTGACTTGAAGTTAACTCAGA 1869
615 pLysLeuProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLysA 632
|||||
1870 TAACTCCCAAGAGATGCTCTCTCTCAGCGCAGCAACGCTGACAGA 1919
632 rGAlaProGlyAspGlnGlyLysTyrIleAspLeuArgHis 646
|||||
1920 GGGCTCCAGAGAGACAGGAGAGAAATACATCGATCTGAGGCAT 1963
seq_name: gb_ro:AB035506

seq_documentation_block:
LOCUS AB035506 2106 bp mRNA linear ROD 03-OCT-2000
DEFINITION Rattus norvegicus mRNA for l-glicerol/MUC18, complete cds.
ACCESSION AB035506
VERSION AB035506.1 GI:10566948
KEYWORDS l-glicerol/MUC18.
SOURCE Rattus norvegicus (strain:Sprague-Dawley) male cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 2106)
AUTHORS Taira,E., Okumura,S. and Miki,N.
TITLE Functional analysis of rat glicerol/MUC18
JOURNLS Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 2106)
AUTHORS Taira,E., Okumura,S. and Miki,N.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Eiichi Taira, Osaka University Medical
School, Department of Pharmacology; 2-2 Yamada-oka, Suita, Osaka
565-0871, Japan (E-mail:etaira@pharm.med.osaka-u.ac.jp,
Tel:+81-6-6879-3521, Fax:+81-6-6879-3521)
FEATURES
source
location/Qualifiers
1..2106
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
17..1963
/gene="l-glicerol/MUC18"
17..1963
/gene="l-glicerol/MUC18"
/codon_start=1
/product="l-glicerol/MUC18"
/protein_id="BAB16048.1"
/db_xref="GI:10566948"
/translation="MGLPRLYCAFLFAACCCRSATGVPGESEKQPTTPDPVEVEGN
TALKCGPAHSGNFSQVEMFLIKERQIPIERFHQKGSQSEPEYERISLHGPGAT
LALSGVPRHDMFLCKSKDRPDHLYOOLYKARPEPTQAMVLGIIHYDIEIKEV
ATCVGRNGYPIPOYIWKNGRPLQDEENRHHIQSQTVESGLITLKSYSARIVKED
KDAQFYCEVSYPSPGNMKSKESEVTVPLVPKAVVEVPEVLKEGQHVAKICLT
DGNOPHPTINKNPSTEMEESTEDENGLSLERPAKHHSYVQCSDLDETITMIS
SDPLFLNVYSVDVDPDTABVEQDSILTLTCAEENODLEFEMLRDKTGOLIGKPS
ILOLNMYREAGRGYICVASYSPGLNRRPRVSGLEGSPMAAKERKYMADENNAL
NLSCBASGHPPTISMNINGSATBMNDPQYVSTLVNLYTPELLERGACATGNSNG
SYTVIILKLVLTITLTHDSSQTTGLSTPYVSIRANSTSTEKKLPQOESKGVYIA
```

```
VIVCTLVAVLAVGATLYFYFKKGLPCGRSGSKOEITLTPTRKSEFVEVKSCLKPEEMA
LLQSGNKGKRAPGQGEKXYDLRH"
BASE COUNT 536 a 573 c 587 g 410 t
ORIGIN
alignment_scores:
Quality: 2587.00 Length: 648
Ratio: 4.297 Gaps: 1
Percent Similarity: 92.901 Percent Identity: 74.537
alignment_block:
US-09-653-961-2 x AB035506 ..
Align seg 1/1 to: AB035506 from: 1 to: 2106
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaIleCysCys 17
|||||
17 ATGGGGCTGCCAGGCGCTGCTGCGCTTCTTATTTGCTGCTGCTGCTG 66
17 SCYSProArgValAlaGlyValProGlyGluAlaGln.....ProA 32
|||||
67 CTGTCCAGCGCCCGCGGTGTGCCAGAGAGAAACAGCCACACCA 116
|||||
32 lAProGluLeuValGluValGluValGlySerThrAlaLeuLysCys 48
|||||
117 CGCCGACCGCGGTGAGGTGAAGTGGCAACACGCGCTTCTGAAGT 166
49 GlyLeuSerGlnSerGlnGlyAsnLeuSerHisValAspIrrPheSerA 65
|||||
167 GGCCTCCACATCCCTCAGCACTTCCAGCAAGTGAAGTGTTTTGAT 216
|||||
65 lHisLysGluLysArgThrLeuIlePheArgValArgGlnGlyGln 82
|||||
217 TCACAGAGAGAGGAGCATACCATTTTCCGTGTGCACAGGAGGAGGCC 266
|||||
82 lInsGluProGlyGluTyrGlnGlnArgLeuSerLeuGlnAspArgGly 98
|||||
267 AGAGTGAACCTGGGGAATGATGACACCCCTCAGCCCTCCAGGCCAGGA 316
|||||
99 AlAthrLeuAlaLeuThrGlnValThrProGlnAspGluArgIlePhe 115
|||||
317 GCTACTGTGGCTGAGTCAAGTACTCTCCATGATATACCGGATGTTCT 366
|||||
115 uCysGlnGlyLysArgProArgSerGlnGluTyrArgIleGlnLeuArg 132
|||||
367 GTTAAGAGCAAGACGACGACGCGCCAGATCACTATGTTCAGTGCAG 416
|||||
132 AlTyrLysAlaProGluGlnProAsnIleGlnValAsnProLeuGlyIle 148
|||||
417 TCTACAAAGCTCCAGAGAGCCAACTATCCAAAGCCAGTGTGGCATC 466
|||||
149 ProValAsnSerLysGluProGluGluValAlaIaThrCysValGlyArg 165
|||||
467 CACGTGGACATACAAAGAGCTTAAGAGGTGCTACGTGTGGGAGGAA 516
|||||
165 nGlyTyrProIleProGlnValIleTyrTyrLysAsnGlyArgProLeu 182
|||||
517 TGGTACCCCATTCCTCAAGTCATCTGTACAAAGAGCGCGGCTTGC 566
|||||
182 YSGlGluLysAsnArgValHisIleGlnSerSerGlnThrValGlnSer 198
|||||
567 AAGAGAGAGAGAACCGCTTATATCCAGTGTACAGACCGGTGATCC 616
|||||
199 SerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValysG 215
|||||
617 AGTGGCTTGTACACTTGAAGAGTGTCTAAGTCCACAGGGTAGCAAGA 666
|||||
215 uAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerG 232
|||||
667 AGACAAAGATGCCCGATTCTACTGTGAAGTCACTACCGCGGCCAGTG 716
|||||
232 lYAsnHisMetLysLeuSerArgGluValThrValProValPheTyrPro 248
```



```

|||||.....
717 GGAACCGCATGAAGGAATCTAAGAGGCTGCTGCTCTTATACCT 766
249 ThrGluValTrrPleuGluValGluProValGlyMetLeuLysGluGlu 265
767 GCAGAAAAAGTGTGGTGGAGGTAGAACCCCTGGGACTGCTGAAGGAGAG 816
265 yAspArgValGluLeuArgCysLeuAlaAspGlyAsnProProHisP 282
817 GGACCATGTGAAGATCAGTGTCTGACTGACGGCAACCCCTCAACCCACT 866
282 heSerLLeSerLysGlnAsnProSerThrArgGluAlaGluGluThr 298
867 TCACATTCACAGAGAAAGAACCCAGCTGAGATGGAAGAGAGAGAGC 916
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
917 ACCGATGAAGAAAGGTCTCTCTCTGAGACCCGCCAGAAAGCACACAG 966
315 rGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetLLeSerLeu 332
967 TGGGCTTACCAATGTCTGAGGCTTGAAGTGAACACTGACTTATGCTGT 1016
332 euSerGluProGlnGluLeuValAsnTyrValSerAspValArgVal 348
1017 CGAGTGACCCCTGGAGCTGTGTGAACTATGTCTGATGATTCAAGTG 1066
349 SerProAlaAlaProGluArgGlnGluGlySerSerLeuThrLeuThrCy 365
1067 GATCCAACTGCCCCCTGAAGCTCCAGAGAGGTATAGCTTCACACTGACCTG 1116
365 sGluAlaGluSerSerGlnAspLeuGluPheGlnTrrPleuArgGluGlu 382
1117 TAAGCAGAGAGATACCGAGACCTTGAGTTGAGTGCTGAGAGCAAGA 1166
382 hrGlyGluValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLys 398
1167 CAGGCACTCTTGTGGAAAGGGGCCCACTTCAAGCTAAACAGCTGAGA 1216
399 ArgGluAlaGlyGlyTyrArgCysValAlaSerValProSerLLePr 415
1217 AGGAGAGAGGGGGAAGATATCTGCTGTGCACTGCTCCAGGCTTCC 1266
415 oGlyLeuAsnArgTrrGlnLeuValAsnValAlaLLePheGlyProT 432
1267 TGGCTTGAATGCACACGCGGGTACGGGTGGCATTTTGGGCCCGCT 1316
432 rPMeAlaPheLysGluArgLysValTrrPValLysGluAsnMetValLeu 448
1317 GGATGGCAGCAAAAGAGAGAGGTGGCGCAAGAAATGCAATGCTG 1366
449 AsnLeuSerCysGluAlaSerGlyHisProArgProThrLLeSerTrrPAs 465
1367 AATCTGTCTTGTGAGGCTTTCAGGCCATCTTCAGCCCACTCTCTGGAA 1416
465 nValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeu 482
1417 TATCAATGGCTCGCACTGAATGAACCTGATCCTCAGACAGTAGAGA 1466
482 erThrLeuAsnValLeuValThrProGluLeuLeuGluThrGlyValGlu 498
1467 GCACCTTAAGTCTTGTGACCCAGAGCTTGGAGACCGGTCAGAGA 1516
499 CysThrAlaSerAsnAspLeuGlyLysAsnThrSerLLeLeuPheLeuGlu 515
1517 TGTACAGGCTCAACTCCCTGGGCTCATACACACAGGCTCATCTTGA 1566
515 uLeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyL 532
1567 GCTGCTCACTTAAACACCTCACACATGACTCCAGCCAAACCACTGGCC 1616
532 euSerThrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThr 548
|||||.....

```

```

1617 TCAGCACCCCAACGATCAGTCTCTACAGCAGAGCCAAACAGACCTCCACA 1666
549 GluArgLysLeuProGluSerArgGlyValValIleValAlaVala 565
1667 GAGAAAAAGTGTCCACAGCAGGAGAGACAAAGTGTGTATCTGTGCTGT 1716
565 lIleValCysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheL 582
1717 GATAGTATGACCTGGTGTCTGCTGACTGGGCGCTACCTCTATTA 1766
582 euTrrLysLysGlyLysLeuProCysArgArgSerGlyLysGlnGlu 598
1767 TCACAGAAAGAGGGCAACCTCTGTGAGAGCTCAGAGAAACAGGAGANC 1816
599 ThrLeuProProSerArgLysSerGluLeuValGluValLysSer 615
1817 ACGGTGCCCCCACTGCTAAGACTGAATTTGTGATTGAAGTTAAGTCACA 1866
615 pLysLeuProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLys 632
1867 TAAGCTCCAGAGAGATGCTCTCTTCAGGGCAGACACGCTGACACAGA 1916
632 rGluAlaProGlyAspGlnGlyGluLysTrrLLeAspLeuArgHis 646
1917 GGGCTCCAGAGAGACAGGAGAGAAATACATCATCTGAGGACT 1960
seq_name: gb_ro:AB035509
seq_documentation_block:
LOCUS AB035509 2772 bp mRNA linear ROD 03-Oct-2000
DEFINITION Mus musculus mRNA for s-glycerin/MUC18, complete cds.
ACCESSION AB035509
VERSION AB035509.1 GI:10566954
KEYWORDS s-glycerin/MUC18.
SOURCE Mus musculus (strain: BALB/c) male cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2772)
AUTHORS Taira, E., Okumura, S. and Miki, N.
TITLE mouse s-glycerin/MUC18
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 2772)
AUTHORS Taira, E., Okumura, S. and Miki, N.
TITLE Direct Submission
JOURNAL Submitted (03-Dec-1999) Eiichi Taira, Osaka University Medical
School, Department of Pharmacology; 2-2 Yamada-oka, Suita, Osaka
565-0871, Japan (E-mail: etaira@pharm.med.osaka-u.ac.jp,
Tel:+81-6-6879-3521, Fax:+81-6-6879-3521)
FEATURES
source
1..2772
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/sex="male"
20..1840
/gene="s-glycerin/MUC18"
20..1840
/gene="s-glycerin/MUC18"
/codon_start=1
/product="s-glycerin/MUC18"
/protein_id="BAB16051.1"
/db_xref="GI:10566954"
/translation="MGLPKYCVFLFAKCCCRRAAGPGEKKOPVPTPDLYAEVGS
TALKGSRASRSGNFSQVDWILIKENQILLFRHOGKGRREGYRHRSLSDSVAT
LALSHVTFHDERMFLCKSKRPRLDQVLEQFAPEPTIQANVGIHVRQELREY
ATCGRWGYPRLQVLAWYKNSLPLOEENRVHIOSSQIVESSGLYTLKSVLSARLEED
KDAQFYCELSYRLPSGNHMKESKEVTVPAKVWVEYVGLKEGDHVTIRCLT
DGNQDPHFTIRKDPSTGEMEEESTDNGLSLEPAKHSHGLXOCOSDLETTILS
SDPLEILVNYSDVQNPATPENOEGSLITTCBAESNODLFEFEMLDKRGQILGKGP
VLQANNRREAGGRYLCKASVPRPVGINRFDOLYSVGIFGSPMAALKRRKRWQENAYL
NLSCASGHDPPTLSMWNYSATETWNPDPQTVSTLVNLVATPELLETGAGCTASNSLG
SNTTIVLKLVTLLTLLIPDSSQITGLSTLIVSPHTRANSISTEKKLPQPSKGVIVA

```

BASE COUNT 675 a 706 c 749 g 642 t
ORIGIN
alignment_scores:
Quality: 2415.50 Length: 648
Ratio: 4.275 Gaps: 2
Percent Similarity: 87.191 Percent Identity: 70.833
alignment_block:
US-09-653-961-2 x AB035509 ..
Align seg 1/1 to: AB035509 from: 1 to: 2772
1 MetGluLeuProArgValAlaGlyValProGluGluAlaCysCysCys 17
|||||
20 ATGGGGCTGCCAAACTGGGTGCTGCTGCTGCTGCTGCTGCTGCTG 69
17 scysProArgValAlaGlyValProGluGluAlaGluGlu.....ProA 32
|||||
70 CTGTGCGCCGCGCGCGGTGTGTGCGAGAGAGAAAGCAGCCAGTACCA 119
32 laProGluLeuValGluValGluValGlySerThrAlaLeuLeuLysCys 48
:|||||
120 CGCCCGACCGTGTGGAGGCGAGAGTGGGCGACAGCCCTTCACAGTGT 169
49 GlyLeuSerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerVa 65
|||||
170 GGGCCCTCAGGGCTCAGGCACTTCAAGCAAGTGGAGCTGGTTTGAT 219
65 lHisLysGluLysArgThrLeuLeuIlePheArgValArgGlnGlnGly 82
:|||||
220 TCACAAAGAGAGGAGACAGTACTGATTTTCGCTGTCGACCAAGCAGGCG 269
82 lInsertLuprogLyuTrgGluGluArgLeuSerLeuLaspArgGly 98
|||
270 AGCGGGAACCTGGTGAATATGAGCACCGGCTTACCTCCAAAGACTCGGTG 319
99 AlaThrLeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLe 115
|||||
320 GCTACTCTGGCCCTGAGTCACTGCTACTCCCATGATGAGCGAATGTCT 369
115 ucysGlnGlyLysArgProArgSerGlnGluTrgArgIleGlnLeuArgV 132
:|||||
370 GTGTAGAGAGCAAGCGACCAAGCGCTCCAGGATCAGTGTGAGCTTCAAG 419
132 alLysrLysAlaProGluLuprogLysValProAsnIleGlnValAsnProLeuGlyIle 148
|||
420 TCTTCAAAAGCCCGAGAGAAACCAATATTCAAGCCAATGTCTGGGGCATC 469
149 ProValAsnSerLysGluLuprogLysValAlaThrCysValGlyTrgAs 165
|||||
470 CATGTGGACAGCGCAAGAGCTCAAGAGGTTCCTACTGTGTGGGGGAA 519
165 nGlyTrpProIleProGlnValIleTrpTrpLysAsnGlyArgProLeuL 182
:|||||
520 CGGCTACCCCATCTCTCAAGTCTATGTATCAAGAACAGTCTGCCCTTGC 569
182 ysgLupLysAsnArgValHisIleGlnSerSerGlnThrValGluSer 198
:|||||
570 AAGAGGAGGAGAAACGATTCAATATCAGCATCACAGATGTGCGAGTCC 619
199 SerGlyLeuTrpThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
|||||
620 AGTGCTGTGACACCTTGAAGAGTCTTGTGAGTGCACCGCTAGATTAGA 669
215 uasPlysAspAlaGlnPheTrpCysGluLeuAsnTrpArgLeuProSerG 232
:|||||
670 AGCAAAAGATGCCAGTTTACTGTGAACCTACCGGCTACCGGCTACGAG 719
232 LysAsnHisMetLysGluSerArgGluValThrValProValPheTrpPro 248
|||||

720 GGAACCATATGAAAGGAATCTAAGAGGTCACTGCTCCTGTTTCTACCT 769
249 ThrGluLysValTrpLeuGluValGluProValGlyMetLeuLysGluG 265
:|||||
770 GCAGAAAAAGTGTGGGTGAGGTAGAGCTTGGGGCTGTGAAGGAAG 819
265 yAspArgValGluIleArgCysLeuAlaAspGlyAsnProProHisP 282
:|||||
820 GGATCATGTGACAAATCAGGTGTGTGACAGATGGCAACCTCAACCCACT 869
282 heserLileSerLysGlnAsnProSerThrArgGluAlaGluGluThr 298
:|||||
870 TCACATATCAACAGAGAGAGCCAGCACTGGGAGATGAGAGAGAGAGC 919
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSe 315
:|||||
920 ACCGATGAATAATGGGCTCTGCTGCTGAGGCTGCCGAAAAGCACATAG 969
315 rGlyArgTrpGluCysGlnGlyLeuAspLeuAspThrMetLileSerLeu 332
|||
970 CGGGCTTACCAAGTGTAGAGTGTGAGCTGGAACATACATCACACTGT 1019
332 euSerGluProGlnGluLeuLeuValAsnTrpValSerAspValArgVal 348
|||
1020 CAAGTGAACCCCTGAGCTGCTGTGAACTATGTGTGATGTTCAAGTG 1069
349 SerProAlaAlaProGluLuprogLysGlnGluLysSerSerLeuThrLeuTrpCys 365
:|||||
1070 AATCCAACTGCCCTCGAAGTCCAGGAAGTGAAGCTCAGCGCTGACTGT 1119
365 sGluAlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGluLut 382
:|||||
1120 CGAGGCAAGAAAGTAAACCGAGCACTTGAATTTGAGTGGCTGAGAACAGA 1169
382 lrrGlyGluValLeuGluLuprogLysProValLeuGlnLeuHisAspLeuLys 398
|||
1170 CAGGCACTGTGTGGAAAGAGGTCCGCTCCAGCTCAAGCTTAACAAAGCTGA 1219
399 ArgGluAlaGlyGlyGlyTrpArgCysValAlaSerValProserIlePr 415
:|||||
1220 CGGGAAGCAGGGGGAGCGTATCTGTGATGTGCAATGTGCCCAAGATGCC 1269
415 oGlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProT 432
:|||||
1270 TGGCTTGAATGTATCCAGCTGCTGAGCTGTGGCATTTTGGGCTCCCAT 1319
432 rPheAlaPheLysGluArgLysValTrpValLysGluAsnMetValLeu 448
|||
1320 GGATGGCATTTAAAGAGAGAGAGGTGTGTGCAAGAGAAATGCAAGTCTGT 1369
449 AsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSerTrpAs 465
:|||||
1370 AATCTGTCTTGAGGCTTCAAGACATCTCAGCCCAACATCTCTCTGAA 1419
465 nValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuS 482
:|||||
1420 TGTCAATGTTCGGCACTGAATGAAGAACCAAGTCCAGACAGTATGTGA 1469
482 erThrLeuAsnValLeuValThrProGluLeuLeuGluThrGlyValGlu 498
:|||||
1470 GCACCTTAATGTCTTGTGACCGCAGAGCTTGTGGAGACAGGTGCGAG 1519
499 CysThrAlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuG 515
:|||||
1520 TGTACAGGCTCAACTCCCTGGGCTCAAAACACACCAACATGTGTTGAA 1569
515 uLeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyL 532
:|||||
1570 GCTGTCACTTTAAACCACTTCAATACCTGACTCCAGCAACCACTGACC 1619
532 euSerThrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThr 548
|||
1620 TCAGCACTTCAAGTCAAGTCTCAGACCAAGCAAGCAAGCACTCCACA 1669

549 GIUArgLysLeuProGluSerArgGlyValValIleValAlaVala 565
|||||
1670 GAGAAAGAGCCACACAGGAGCAAGAGTGGTGCATCTGGCTGT 1719
565 ILeValCysIleLeuValIleValAlaValIleValLeuTyrPheL 582
|||||
1720 CATATGTGTACCTTGTGCTGTGCTGCTGCTGCTGCTGCTATTTCC 1769
582 eUTYrLysLysGlyLysLeuProCysArgArgSerGlyLysGlnGluIle 598
|||||
1770 TCTACAAAGAGGCAAGCTGCATGTGAGCGTGGGAAACAGAGAT. 1818
599 ThrLeuProProSerArgLysSerGluLeuValValGluValLysSerAs 615
1818 1818
615 PLYSLeuProGluGluMetGlyLeuGlnGlySerSerGlyAspLysA 632
1818 1818
632 rGAlAProGlyAspGlnGlyLysTyrIleAspLeuArgHis 646
1819GGAGGAAATACATCATGATCTGAGGCAT 1845
seq_name: gb_ro:AB035507
seq_documentation_block: 1988 bp mRNA linear ROD 03-OCT-2000.
LOCUS AB035507 1988 bp mRNA linear ROD 03-OCT-2000.
DEFINITION Rattus norvegicus mRNA for s-glicerol/MUC18, complete cds.
ACCESSION AB035507
VERSION AB035507.1 GI:10566950
KEYWORDS s-glicerol/MUC18
SOURCE Rattus norvegicus (strain: Sprague-Dawley) male heart cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1988)
AUTHORS Taira, E., Okumura, S. and Miki, N.
JOURNAL rat s-glicerol/MUC18
PUBLISHED ONLY in Database (2000) In press
REFERENCE 2 (bases 1 to 1988)
AUTHORS Taira, E., Okumura, S. and Miki, N.
JOURNAL Direct Submission
SUBMITTED (03-DEC-1999) Eiichi Taira, Osaka University Medical
School, Department of Pharmacology; 2-2 Yamada-oka, Suita, Osaka
565-0871, Japan (E-mail: etaira@pharm.med.osaka-u.ac.jp,
Tel:+81-6-6879-3521, Fax:+81-6-6879-3521)
FEATURES
source
1. 1988
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="heart"
17. 1837
/gene="s-glicerol/MUC18"
17. 1837
/gene="s-glicerol/MUC18"
/codon_start=1
/product="s-glicerol/MUC18"
/protein_id="PAB16049.1"
/db_xref="GI:10566951"
/translation="MGIPRLVCAFLFAACCCSRAGVPEEKKPTPTPPVVEVKN
TALKCGPAHSGNFSQVEMFLHKEKQIPTRFHQSGSEFEBHRLSLGPGAT
LALSQVPHDRFLKSKQRPQDHYVQLOVYKAPPEPTIQANVLGIHVDIQLKEY
ATCYGRNGYIPQVITWYKNGRPLQEEENRYVLIOSQVSESSGLYTLKSYLSARVVED
KDAQFYEVSYRPPSGNRKESKEVTVLVPKAVVEYEPGLKEDGDKVTRICT
DGNQPHPTINKKSPTEEMEESESTDENGLLEPACKHSGYVCOOSDLETTVMIS
SDPLELIVNYSDVQVPTAPRYQEGBSLITLTKAESNOLDEBEMLEDRKQGLGKGP
ILQLNNVREKGRITLCVASPSVPGINRTRVSVGLFSGPMNAKARKKWAQENAL
NLSCASGHDPITSMNINGSATEMNDPQTVSTLNLVTPLELLETGACBTASNSIG

SYTIVILIKITLTTLTHDSQTTGLSTPTVPSHRANSTSTKRLPQDSKGVIVA
VIVCVLIVAVGATLYLFFYKKGKRLPCGRSGKQEMERTSI"
BASE COUNT 503 a 544 c 553 g 388 t
ORIGIN
alignment_scores:
Quality: 2360.50 Length: 648
Ratio: 4.193 Gaps: 2
Percent Similarity: 86.883 Percent Identity: 68.981
alignment_block:
US-09-653-961-2 x AB035507 ..
Align seg 1/1 to: AB035507 from: 1 to: 1988
1 MetGlyLeuProArgLeuValCysAlaPheLeuAlaLacCysCys 17
|||||
17 ATGGGCGTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 66
17 sCysProArgValAlaGlyValProGlyGluAlaGln.....ProA 32
|||||
67 CTGTCCAGCGCCAGCGGTGTGCGAGAGAGAAAGCACGCCACACCA 116
117 CGCGGACCGCGGTGAGGTAGAAAGTGCGCAACAGCGCCCTTCGAAGT 166
49 GlyLeuSerGlnSerGlnGlyAsnLeuSerHisValAspTyrPheSer 65
|||||
167 GCGCCCGCACATCCCTCAGCAGCACTCAGCAAGTGGAATGTTTGTGAT 216
65 IHisLysGlyLysArgThrLeuPheArgValArgGlnGlyGlnGly 82
|||||
217 TCACAAAGAGAGGAGATACCATTTTCCGTGTCACACAGGAGAGGCC 266
82 InSerGluProGlyGlyTyrGlnGlnArgLeuSerLeuGlnAspArgGly 98
|||||
267 AGAGTGAACCTGGGGAATATGAACACCGCTCAGCCGCCAGCGCCAGCA 316
99 AlaThrLeuAlaLeuThrGlnValThrProGlnAspGlnArgLLePheLe 115
|||||
317 GCTACTCTGGCCCTGAGTCAAGTCACTCCCATGATGATGCGGATGTTCC 366
115 uCysGlnGlyLysArgProArgSerGlnGlyTyrArgGlnLeuArg 132
|||||
367 GTTAAGAGCAGCAGCAGCAGCGCCCGAGATCACTATGTTCTACCTGCA 416
132 aTYrLysAlaProGluGluProAsnIleGlnValAsnProLeuGlyIle 148
|||||
417 TCTACAAAGGCTCCAGAGAGAGCAACTATCCAAAGCAACGCTGGGCATC 466
149 ProValAsnSerLysGluProGluGlnValAlaThrCysValGlyArgAs 165
|||||
467 CACGTGACATCAAGAGCTTAAAGAGAGTGTACTGTTGGGAGAGAA 516
165 nGlyTyrProIleProGlnValIleTyrLysAsnGlyArgProLeuL 182
|||||
517 TGGCTACCCCATTCCTCAAGTCATCTGTACAAAGAGCGCGGCCCTTCC 566
182 ySGluGluLysAsnArgValHisIleGlnSerSerGlnThrValGluSer 198
|||||
567 AAGAGAGAGAGAACCGCGTTCATATCCAGTGTCCACAGACCGTGGAGTCC 616
199 SerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaIleuValLysGln 215
|||||
617 AGTGGCTTGTACACCTTGAAGAGTGTCTAAGTGCACGGGTAGTCAAGGA 666
215 uAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 232
|||||
667 AGACAAAGATGCCAGATTCTTACTGTAAGTACAGTACCGGCGGCAGCTG 716
232 IAsnHisMetLysGluSerArgGluValThrValProValPheTyrPro 248

```

|||||.....
717 GGAAACGCGAAGAGATCTAAGAGAGTCACTGCTCCCTGCTTATACCT 766
249 ThGluTyValTrPleuGluValGluProValGluMetLeuLysGlu 265
767 GCGAAAAAGCTGGGTGGAGGTAGAACCCGTGGACTGTGAGGAAG 816
265 yAspArgValGluLeuArgCysLeuAlaAspGlyAsnProProPhe 282
817 GGACCATGTGATGATCAGGTGTGTGACTGACGGCAACCTCAACCCACT 866
282 heSerLysGlnAsnProSerThrArgGluAlaGluGluThr 298
867 TCACATATCAACAGAAAGAACCCAGACCTGAGAGATGAGAGGGAGAC 916
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGlu 315
917 ACCGATGAAACGCTCTCTCTGAGACCCGCCAGAACGACACAG 966
315 rGlyArgTyGluCysGlnGlyLeuAspLeuAspThrMetLysLeu 332
967 TGGGCTTACCATGTACAGAGCTGACCTGGAACCTACAGATTATGCTGT 1016
332 euSerGluProGlnGluLeuValAsnTyValSerAspAlaArgVal 348
1017 CGAGTAGCCCTGAGCTGTGTGTAATGTGTGTATGTTCACAGTG 1066
349 SerProAlaAlaProGluArgGlnGluGlySerSerLeuThrLeu 365
1067 GATCCAACTGCCCTGATAGTCCAGAGAGTGAAGCTTACACTGACCTG 1116
365 sGluAlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGlu 382
1117 TTAGCGAGAGAGTACACAGACCTGAGTTGAGTGGCTGAGAGACAGA 1166
382 hGlyGlnValLeuGluArgGlyProValLeuGlnLeuLysLeu 398
1167 CAGGCGAGCTCTGGGAAAGGGGCCATCTTCAGTAAACAACGGGAA 1216
399 ArgGluAlaGlyGlyTyArgCysValAlaSerValProSerLys 415
1217 AGGAGACAGGGGAGATATCTGTGTGTCATGTCCAGAGGTTCC 1266
415 oGlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyPro 432
1267 TGGCTTGAATCCACACAGCGGCTGAGCGTGGCATTTTGGGTCCCT 1316
432 rPheAlaPheLysGluArgLysValTrpValLysGluAsnMetVal 448
1317 GATGGCACCAAGAGAGAGAGGTGTGGCCAGAGATCAATGCTG 1366
449 AsnLeuSerCysGluAlaSerGlyAsnProArgProThrLysSer 465
1367 AATGTGTCTTGAGGCTTCAGGCCATCTCAGCCACCATCTCTCGGAA 1416
465 nValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArgVal 482
1417 TATCAATGGCTCGCAACATGAATGGAACCTCATCTCAGACATGATGA 1466
482 eThrLeuAsnValLeuValTrpGluLeuLeuGluThrGlyValGlu 498
1467 GCACCTTGATGTCTGTGACCCAGAGCTTCTGGAGACCGGTGCAGAA 1516
499 CysThrAlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPhe 515
1517 TGTACAGCTCAACTCCTGTGGCTCATACACAGGTCATCATCTGAA 1566
515 uLeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrGly 532
1567 GCTGTGACATTTAACCACTTCACATGATGACAGCAACCACTGGCC 1616
532 euSerThrSerThrAlaSerProHisThrArgAlaAsnSerThrSer 548
|||||.....

```

```

1617 TCAGACCCCCACAGTCAGTCCCTCACAGACGCAACAGCACCTCCACA 1666
549 GluArgLysLeuProGluProGluSerArgGlyValValIleValAla 565
1667 GAGAAAAAGCTGCCACAGAGAGCAAGAGTGGTGCATCGTGGCTGT 1716
565 lIleValCysIleLeuValLeuAlaValLeuGlyAlaValLeuTyPhe 582
1717 GATATATGTAGCGCTGTGCTGTGCTGCTGAGGCTTACCTCATACT 1766
582 euTyTrpLysGlyLysLeuProCysArgArgSerGlyLysGlnIle 598
1767 TCTACAAAGAGCGCAACTGCTGTGGACGCTCAGGAAAAACAGAGAT 1815
599 ThrLeuProProSerArgLysSerGluLeuValValGluValLysSer 615
1815 ..... 1815
615 pLysLeuProGluGluMetGlyLeuLeuGlnGlySerSerGlyAsp 632
1815 ..... 1815
632 rGAlaProGlyAspGlnGlyLysTyTrpLeuAspLeuArgHis 646
1816 .....GGAGAGAAATACATGATGTGAGGCAT 1842
seq_name: gb_hhg:AP001557
seq_documentation_block:
LOCUS AP001557 182429 bp DNA linear HTG 03-NOV-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-680A7 map 11q23, WORKING
DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION AP001557.3 GI:11094164
VERSION AP001557
KEYWORDS HTG, HTGS, PHASRL, HTGS, DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-680A7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 182429)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 182,429 genomic DNA of 11q23
Published Only in Database (2000) In press
2 (bases 1 to 182429)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
JOURNAL Submitted (29-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hhg.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Nov 3, 2000 this sequence version replaced gi:8117391.
COMMENT
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hhg.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDrafl11
Center clone name: RP11-680A7
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 175597 bases at least Q40
Consensus quality: 178555 bases at least Q30
Insert size: 180429; sum-of-ctrls
Quality coverage: 9.01x in Q20 bases; sum-of-ctrls

```

NOTE: This is a 'working draft' sequence. It currently consists of

21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 30107 contig of 30107 bp in length
30208 54357 contig of 24150 bp in length
54458 69839 contig of 15382 bp in length
69940 84048 contig of 14109 bp in length
84149 95177 contig of 11029 bp in length
95278 10763 contig of 12486 bp in length
107864 119057 contig of 11194 bp in length
119158 128134 contig of 8697 bp in length
128235 136903 contig of 8864 bp in length
137004 145887 contig of 8884 bp in length
145988 153199 contig of 7212 bp in length
153300 159565 contig of 6266 bp in length
159666 163580 contig of 3915 bp in length
163681 165152 contig of 1472 bp in length
165253 167001 contig of 1649 bp in length
167002 170460 contig of 3459 bp in length
170461 172761 contig of 2201 bp in length
172762 172861 contig of 100 bp in length
172862 175605 contig of 2744 bp in length
175606 175705 contig of 100 bp in length
175706 178308 contig of 2603 bp in length
178309 178408 contig of 100 bp in length
```

NOTE: This is a "working draft" sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 30107 contig of 30107 bp in length
30108 30207 contig of 100 bp in length
30208 54357 contig of 24150 bp in length
54358 54457 contig of 100 bp in length
54458 69839 contig of 15382 bp in length
69840 69939 contig of 100 bp in length
69940 84048 contig of 14109 bp in length
84049 84148 contig of 100 bp in length
84149 95177 contig of 11029 bp in length
95178 95277 contig of 12486 bp in length
95278 10763 contig of 100 bp in length
10764 107863 contig of 11194 bp in length
107864 119057 contig of 11194 bp in length
119058 119157 contig of 8697 bp in length
119158 128134 contig of 8697 bp in length
128135 128234 contig of 100 bp in length
128235 136903 contig of 8669 bp in length
136904 137003 contig of 100 bp in length
137004 145887 contig of 8884 bp in length
145888 145987 contig of 100 bp in length
145988 153199 contig of 7212 bp in length
153200 153299 contig of 100 bp in length
153300 159565 contig of 6266 bp in length
159566 159665 contig of 100 bp in length
159666 163580 contig of 3915 bp in length
163581 163680 contig of 1472 bp in length
163681 165152 contig of 1649 bp in length
165153 165252 contig of 100 bp in length
165253 166901 contig of 1649 bp in length
166902 167001 contig of 100 bp in length
167002 170460 contig of 3459 bp in length
170461 172761 contig of 2201 bp in length
172762 172861 contig of 100 bp in length
172862 175605 contig of 2744 bp in length
175606 175705 contig of 100 bp in length
175706 178308 contig of 2603 bp in length
178309 178408 contig of 100 bp in length
```

FEATURES

source

1. 182429

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11q23"

/clone="RP11-680A7"

1. 30107

/note="assembly-fragment"

30208. 54357

/note="assembly-fragment"

54458. 69839

/note="assembly-fragment"

69940. 84048

/note="assembly-fragment"

84149. 95177

/note="assembly-fragment"

95278. 10763

/note="assembly-fragment"

10764. 119057

/note="assembly-fragment"

119158. 128134

/note="assembly-fragment"

128235. 136903

/note="assembly-fragment"

137004. 145887

/note="assembly-fragment"

145988. 153199

/note="assembly-fragment"

153300. 159565

/note="assembly-fragment"

159666. 163580

/note="assembly-fragment"

163681. 165152

/note="assembly-fragment clone_end:T7 vector_side:left"

165253. 166901

/note="assembly-fragment clone_end:SP6 vector_side:left"

167002. 170460

/note="assembly-fragment"

170561. 172761

/note="assembly-fragment"

172862. 175605

/note="assembly-fragment"

175706. 178308

/note="assembly-fragment"

178409. 180918

/note="assembly-fragment"

181019. 182429

/note="assembly-fragment"

BASE COUNT 45724 a 42083 c 43875 g 48747 t 2000 others

ORIGIN

alignment_scores: Length: 870

Ratio: 4.264 Gaps: 15

Percent Similarity: 51.379 Percent Identity: 51.264

alignment_block:

us-09-653-961-2 x AP001557/rev ..

Align seg 1/1 to reverse of: AP001557 from: 1 to: 182429

188 ValHisIleGlnSerGlnThrValGluSerSerGlyLeuTyrThrle 204

29642 GTCCACATTCAGTGTCCACAGCTGTGAGTCCAGTGCTTGTACACCTT 29593

204 uGlnSerIleLeuTysAlaGlnLeuValTysGlyAspLysASPAlaGlnP 221

29592 GCAGAGTATTCGAGGACAGCTGTGAAGAGCAAGACAAAGATGCCAGT 29543

```
221 heTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnHisMetLysGlu 237
|||||
29542 TTACTGTGAGCTCACTACCGGCTGCCCTGGGAGAACCATGAAAGAG 29493
238 SerArgGluValThrValProValPhe..Tyr..... 247
|||||
29492 TCCAGGGAAGTCACTCCCTGTTTCTGTGTGAGTACTAACCTGACTCTCT 29443
247 ..... 247
29442 GCTTACTGGGCTGGGCTGCTGTGAAGTCTGCTGCTCCCTTATCC 29393
247 ..... 247
29392 AGCCGAGGTTGACTCATGGGCCAAGAGCTACCCCTTTGCTATGCAGA 29343
248 .ProThrGluLysValTrpLeuGluValGluProValGlyMetLeuLysG 264
|||||
29342 CCCGACAGAAAAAGTGTGGCTGGAAGTGAAGCCCGTGGGAATGCTGAAGG 29293
264 LuGlyAspArgValGluIleArgCysLeuAlaAspGlyAsnProProPro 280
|||||
29292 AAGGGGACCGCGTGGAAATCAGGTGTTGCTGATGCAACCTCCACCA 29243
281 HisPheSerIleSerLys..... 286
|||||
29242 CACTTCAGCATCAGCAAGCAGGTGTGGGGCTTTGTGCTGGTGCAAGTG 29193
286 ..... 286
29192 GGGATGCAGACAGAGACCCAGGTGAGTGGCGGAGGTGCACCTCTCTCA 29143
287 .....Gln..AsnProSerThrArgGluAlaGluGluLeuT 298
|||||
29142 CTTCCCAATCTCTCCCTAGAACCCAGCACAGGAGGAGCAAGAGAGA 29093
298 hrThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 314
|||||
29092 CAACCAACGCAACAGGGGTCTGTGTCTGTGAGCTGCCCGGAGAGAAAC 29043
315 SerGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLe 331
|||||
29042 AGTGGGCGCTATGAAATGTCAGGGCCCTGGACTGTGACACCATGATATCGCT 28993
331 uLeuSerGluProGlnGluLeuLeuValAsn..Tyr..... 342
|||||
28992 GCTGAGTGAACCAAGAACTACTGTGACTGTGAGGGGCTGGGAGACC 28943
342 ..... 342
28942 AGGACAGGGGACACAGCTGGGGCAATAGGGAAGTGGCCCCACCTGCC 28893
343 .....ValSerAspValArgValSerP 350
|||||
28892 TGGGTGACATGATGCTGTAAACCCAGATGTCTCTGACGTCCGAGTGAGTC 28843
350 roAla..AlaProGluArgGlnGlyLysSerSerLeuThrLeuThrCysG 366
|||||
28842 CCGGAGGCGCTGTAGAGACAGGAAGCAGACGCTCACCTGACCTGTGTA 28793
366 uAlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGluGlu.... 381
|||||
28792 GGCAGAGAGTGAAGCAAGACCTCGAGTTCAGTGGCTGAGGAAGAGGTAC 28743
381 ..... 381
28742 CCAAGCGGCTCTGGGCGCAAGGGCTGATGGGTGGGTGGGCGTGGCGGTG 28693
382 .....Th 382
28692 ACAAGAATGTGTCTTTCCAGACCTCTGACCATCTCTATCCACAGAC 28643
|||||

382 rGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysA 399
|||||
28642 AGGCCAGGTGCTGGAAAGGGGCTGTGCTTCAAGTTGATGACCTGAAAC 28593
399 rGlyAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
|||||
28592 GGGAGCAGAGAGCGGCTATGCTCTGCTGGCTGTGTGCTCCAGCATACCC 28543
416 GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePhe..... 428
|||||
28542 GGCCTGAACCGCACACAGCTGTCAACGTGSCATTTTGTGTGAGGCGCT 28493
428 ..... 428
28492 CCTCTGGGTAGAGACCAAGCTACCCCAAGTGGGGTGTTTTAAAGCTCT 28443
428 ..... 428
28442 TTGACAAAAAGCCACCTGCTGCTGGGAGCTGTGTGGCGAGGGGAG 28393
429 .....GlyPro 430
28392 GCAGGCTAGCTGGGCTTGCCACTCACTCTCTCTACCTCCCCAGAGGCC 28343
430 oProThrMetAlaPheLysGluArgLysValTrpValLysGluAsnMetV 447
|||||
28342 CCTTTGATGGCATTCAGAGAGAGAAAGTGTGGTGAAGAAGCAATATAG 28293
447 aLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSer 463
|||||
28292 TGTGATATGTCTGTGTGAAGGCTCAGGGCACCCCGGCCACATCTCC 28243
28242 TGAAGGTCAACGCAAGGTGAGCGGTGTCTCATCTGCACACTCTCTCT 28193
464 TrpAsnValAsnGlyThr..... 469
|||||
469 ..... 469
28192 AATCCTGGCTGGCTTTTGTGGCAGACAGATCTCCCATCCCT 28143
470 .....AlaSerGluGlnAspGlnAspPro 477
|||||
28142 GAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 28093
478 GlnArgValLeuSerThrLeuAsnValLeuValThrProGluLeuGlu 494
|||||
28092 CAGCGAGTCTGAGCACTCTGAAATGCTCTGTGACCCCGGAGCTGTGGA 28043
494 uThrGlyValGluCysThrAlaSerAsnAspLeuGlyLysAsnThrSerI 511
|||||
28042 GACAGGTGTGAATGACAGGGCTCCACAGCACTGGGCAAAAACACCA 27993
511 LeLeuPheLeuGluLeu..Val..... 517
|||||
27992 TCCTTCTCTGAGACTGGGTGAGGGCTGCATCCCTGAGGGATGACAGG 27943
517 ..... 517
27942 GCGAGGCCAGGACAGGGGAGTAATTCGCTCTCTCTTAACACCTCC 27893
518 .....AsnLeuThrThrLeuThrProAspSerAsnThrThr 529
|||||
27892 TCTGCTCTCTAATTCATTAACCACTCTACACCAAGATCCAAACACAC 27843
530 ThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaAsnSerTh 546
|||||
27842 ACTGGCTCTGACACTTCACCTGCACTCTCTATACCAAGCCAAAGAC 27793
546 rSerThr..... 548
|||||
27792 CTCACAGGTAAGCCAGGCTGGCAAGAACAGAGGCTGTGCCAAGGCAT 27743
548 ..... 548
```

27742 CCTTCTGCCCTGCTCCCTCCAGAGAGCCCTGTCAGAAAGGTGATAG 27693
 548 548
 27692 CAGCCCATCTTGTGGGCTGAGACTGGGCGAAGCATGTGACGAA 27643
 549 549
 27642 GTGGCTGGGCGAGGAGTACAGAGAGCTCTTTGTGGCGGAGAGAGAA 27553
 551 ysluProglupProgluserAArglyVal11leValAlaVal11leVal 567
 27592 AGCTGCGGAGCGGAGAGCGGGCGGCGGTCATCGTGTGATTGTG 27543
 568 Cys11leVal11leValAlaVal11leVal11leValPheLeuTyrLy 584
 27542 TGCATCTGCTGCTGCGGCTGCTGCGGCTGCTCTCTATTTCTCTATA 27493
 584 slysglysluProgluserAArglyVal11leVal11leVal11leVal 598
 27492 GAAGGCGAGAGCTGCTGCGGCTGCTGCGGAGAGAGATGATGAGCC 27443
 598 598
 27442 AACACGCGGCGCTGCGGCGGCGGCTGCTGCTGCTGCTGCTGCT 27393
 598 598
 27392 GCCCAGGCGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 27343
 598 598
 27342 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27293
 598 598
 27292 TCATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27243
 598 598
 27242 TCATTCACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27193
 599 599
 27192 TACCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27143
 604 glysserGlyLeuVal11leVal11leVal11leVal11leVal11leVal 621
 27142 TAAGAGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27093
 621 etgLyLeuLeuGlnGlySerGlyAspLyAspLyAlaProglAspGln 637
 27092 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27043
 seq_name: gb_htg:AP001003
 seq_documentation_block: 173038 bp DNA linear HTG 30-MAY-2000
 LOCUS AP001003 Homo sapiens chromosome 11 clone RP11-680C5 map 11q23, WORKING
 DEFINITION DRAFT SEQUENCE, 28 unordered pieces.
 ACCESSION AP001003.2 GI:8117674
 VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Homo sapiens DNA, clone:RP11-680C5.
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Authors
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE
 Homo sapiens 173,038 genomic DNA of 11q23
 JOURNAL
 Published Only in Database (2000) In press

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

2 (bases 1 to 173038)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail:hattori@gsc.riken.go.jp;
 URL: <http://hgp.gsc.riken.go.jp/>,
 Fax: 81-42-778-9924)
 On May 31, 2000 this sequence version replaced gi:6997762.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: <http://hgp.gsc.riken.go.jp/>
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraT11
 Center clone name: RP11-680C5
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 149676 bases at least Q40
 Consensus quality: 160868 bases at least Q30
 Consensus quality: 167161 bases at least Q20
 Insert size: 170338; sum-of-ctrls
 Quality coverage: 4.33x in Q20 bases; sum-of-ctrls

 NOTE: This is a 'working draft' sequence. It currently consists of
 28 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved
 1 15558 contig of 15558 bp in length
 15659 33689 contig of 18031 bp in length
 33790 47432 contig of 13643 bp in length
 47533 63392 contig of 15860 bp in length
 63493 74234 contig of 10742 bp in length
 74335 85357 contig of 11023 bp in length
 85458 93597 contig of 8140 bp in length
 93698 101983 contig of 8286 bp in length
 102084 110022 contig of 7939 bp in length
 110123 118116 contig of 7994 bp in length
 118217 126464 contig of 8248 bp in length
 126565 132136 contig of 5572 bp in length
 132237 135290 contig of 3054 bp in length
 135391 140229 contig of 4839 bp in length
 140330 143896 contig of 3567 bp in length
 143997 145752 contig of 3576 bp in length
 147673 150446 contig of 2774 bp in length
 150547 153715 contig of 3169 bp in length
 153816 157273 contig of 3458 bp in length
 157374 159817 contig of 2444 bp in length
 159918 161743 contig of 1826 bp in length
 161844 163684 contig of 1432 bp in length
 163785 166657 contig of 1441 bp in length
 166317 168307 contig of 1150 bp in length
 168758 170030 contig of 1823 bp in length
 168408 171725 contig of 1395 bp in length
 170131 173038 contig of 1213 bp in length
 171826
 Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 15558: contig of 15558 bp in length
* 1559 15658: gap of 100 bp
* 33689 33689: contig of 18031 bp in length
* 33690 33789: gap of 100 bp
* 47433 47433: contig of 13643 bp in length
* 47533 47533: gap of 100 bp
* 63393 63492: contig of 15860 bp in length
* 63393 63492: gap of 100 bp
* 74233 74334: gap of 100 bp
* 74335 85357: contig of 11023 bp in length
* 85358 85457: gap of 100 bp
* 93598 93697: gap of 100 bp
* 93598 93697: contig of 8140 bp in length
* 101984 102083: gap of 100 bp
* 102084 110122: contig of 7939 bp in length
* 11023 118116: contig of 7994 bp in length
* 118117 118216: gap of 100 bp
* 118217 126464: contig of 8248 bp in length
* 126465 126564: gap of 100 bp
* 126565 132136: contig of 5572 bp in length
* 132137 132236: gap of 100 bp
* 132237 135290: contig of 3054 bp in length
* 135291 135390: gap of 100 bp
* 135391 140229: contig of 4839 bp in length
* 140330 140330: gap of 100 bp
* 140330 143896: contig of 3567 bp in length
* 143897 143996: gap of 100 bp
* 143997 147572: contig of 3576 bp in length
* 147573 147672: gap of 100 bp
* 147673 150446: contig of 2774 bp in length
* 150447 150546: gap of 100 bp
* 150547 153715: contig of 3169 bp in length
* 153716 153815: gap of 100 bp
* 153816 157273: contig of 3458 bp in length
* 157274 157373: gap of 100 bp
* 157374 159817: contig of 2444 bp in length
* 159818 159917: gap of 100 bp
* 159918 161743: contig of 1826 bp in length
* 161744 161843: gap of 100 bp
* 161844 163684: contig of 1841 bp in length
* 163685 163784: gap of 100 bp
* 163785 165216: contig of 1432 bp in length
* 165217 165316: gap of 100 bp
* 165317 166657: contig of 1341 bp in length
* 166658 166757: gap of 100 bp
* 166758 168307: contig of 1550 bp in length
* 168308 168407: gap of 100 bp
* 168408 170030: contig of 1633 bp in length
* 170031 170130: gap of 100 bp
* 170131 171725: contig of 1595 bp in length
* 171726 171825: gap of 100 bp
* 171826 173038: contig of 1213 bp in length.

FEATURES
    source
        1..173038
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /map="11q23"
            /clone="RP11-680C5"

misc_feature
    1..15558
        /note="assembly_fragment"

misc_feature
    15659..33689
        /note="assembly_fragment"

misc_feature
    33790..47432
        /note="assembly_fragment"

misc_feature
    47533..63392
        /note="assembly_fragment"

misc_feature
    63493..74234
        /note="assembly_fragment"

misc_feature
    74335..85357
        /note="assembly_fragment"

misc_feature
    85458..93597
        /note="assembly_fragment"

misc_feature
    93698..101983
        /note="assembly_fragment"

misc_feature
    102084..110122
        /note="assembly_fragment"

misc_feature
    11023..118116
        /note="assembly_fragment"

misc_feature
    118117..118216
        /note="assembly_fragment"

misc_feature
    118217..126464
        /note="assembly_fragment"

misc_feature
    126465..126564
        /note="assembly_fragment"

misc_feature
    126565..132136
        /note="assembly_fragment"

misc_feature
    132137..132236
        /note="assembly_fragment"

misc_feature
    132237..135290
        /note="assembly_fragment"

misc_feature
    135291..135390
        /note="assembly_fragment"

misc_feature
    135391..140229
        /note="assembly_fragment"

misc_feature
    140330..140330
        /note="assembly_fragment"

misc_feature
    140330..143896
        /note="assembly_fragment"

misc_feature
    143897..143996
        /note="assembly_fragment"

misc_feature
    143997..147572
        /note="assembly_fragment"

misc_feature
    147573..147672
        /note="assembly_fragment"

misc_feature
    147673..150446
        /note="assembly_fragment"

misc_feature
    150447..150546
        /note="assembly_fragment"

misc_feature
    150547..153715
        /note="assembly_fragment"

misc_feature
    153716..153815
        /note="assembly_fragment"

misc_feature
    153816..157273
        /note="assembly_fragment"

misc_feature
    157274..157373
        /note="assembly_fragment"

misc_feature
    157374..159817
        /note="assembly_fragment"

misc_feature
    159818..159917
        /note="assembly_fragment"

misc_feature
    159918..161743
        /note="assembly_fragment"

misc_feature
    161744..161843
        /note="assembly_fragment"

misc_feature
    161844..163684
        /note="assembly_fragment"

misc_feature
    163685..163784
        /note="assembly_fragment"

misc_feature
    163785..165216
        /note="assembly_fragment"

misc_feature
    165217..165316
        /note="assembly_fragment"

misc_feature
    165317..166657
        /note="assembly_fragment"

misc_feature
    166658..168307
        /note="assembly_fragment"

misc_feature
    168308..170030
        /note="assembly_fragment"

misc_feature
    170031..170130
        /note="assembly_fragment"

misc_feature
    170131..171725
        /note="assembly_fragment"

misc_feature
    171726..171825
        /note="assembly_fragment"

misc_feature
    171826..173038
        /note="assembly_fragment"

alignment_scores:
    quality: 1793.00      Length: 872
    Ratio: 4.066          Gaps: 19
    Percent Similarity: 50.573   Percent Identity: 50.344

Alignment_block:
US-09-653-961-2 x AP001003/reverse ..

Align seg 1/1 to reverse of: AP001003 from: 1 to: 173038

188 valhisieginserserginrtvralgisuserserlgyleutyrrhie 204
|||||
33457 GTCCACATTCAGTGTGCTCCACACTGGAGTGCGATTGTGTACACTT 33408
|||||

204 uclnserlileuleylsalaclnleuvallysgluasplysasplaglap 221
|||||
33407 GCAGAGTAATCTCTGAAGGCACGACGTGGTTAAAGAACAAAGATGCCAGT 33358
|||||

221 heTtyCySGluLeuaSnTyFaiRgleProSeGlyAsNHsmetyLsgLu 237
|||||
33357 TTACTAGTAGACTCACTACCGGCTCCACAGTGGAGAACACATGAGAAG 33308
|||||

238 seArGgIuValThrVaIProValPe..... 246
|||||
33307 TCCAGGAAAGTCACGTCCTGTTTTCTGTAGTACTAAGTACTGCTCTC 33258
|||||

246 ..... 246
33257 GCTTACTGGGCTGGGCTGCTGGAAACTCTGCTGCTCCCCCTATATCA 33208
|||||
247 .....Tyr..P 248
33207 GCCCAGGTGTACATCATGGCCAAGCATCCCTTTGCTGATGACAGACC 33158

```



```
248  rorthGluValTrpLeuGluValGluProValGlyMetLeuLysGlu 264
|||||
33157  CGACGAAAAAGTGTGGCTGGAGTGGAGCCCTGGGGAATGCTGAAGAA 33108
265  GlyAspArgValGluIleArgCysLeuAlaSpGlyAsnProProProH 281
|||||
33107  GGGGACCGCGTGGGAATCAGGTGTGGCTGATGACCACTCCACACACA 33058
281  sPheSerIleSerLys..... 286
|||||
33057  CTTGAGCATCAGCAAGCAGGTGTGGGGCTTTGTCTGGTGCAGTGGG 33008
286  ..... 286
33007  GATGCACAGACAGACCAGGTGGAGTGGGGAGTGCATCTGTCTCAC 32958
287  .....Gln..AsnProSerThrArgGluAlaGluGluThr 298
|||||
32957  TCCCAATCTCTCTAGAACCCAGCACACAGGAGCAGAGAGAGACA 32908
299  ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
|||||
32907  ACCAACGACACAGGGGTCTGTGCTGAGGCTGCCGGAAGAACACAG 32858
315  rGlyArgTyrGlyCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 332
|||||
32857  TGGGGCTATGATGTCCAGGGCTTGACTGTGACACATGATATGCTGC 32808
332  eusSerGluProGlnGluLeuValAsn..... 341
|||||
32807  TGAGTGAACACAGGAACTACTGTGTAAGTGTGAGGGGCTGGGACCC 32758
341  ..... 341
32757  GACAGGGGACAGAGCTGGGGCAATAGGGAAGTGGCCCACTGCTGCT 32708
342  .....Tyr..ValSerAspValArgValSerPr 350
|||||
32707  GGTGACATGATGTCTGTACCCCAAGATGTGTGACGTCCGAGTGATGC 32658
350  oAlaAlaProGluArgGlnGluLysSerSerLeuThrLeuThrCysGlu 366
|||||
32657  CGCAGCCCTGAGAGACAGGAAGGAGCAGCCCTGACCTGACCT..GTGA 32611
367  AlaGluSerSerGlnAspLeuGluPheGlnThrLeuArgGlu..Glu.... 381
|||||
32610  GCAGAGAGTAGCCAGAGGCTCGAGTTCCAGAGGCTGAGAGAGAGGTACC 32561
381  ..... 381
32560  CAGGGGGGTCTGGGGCAAGGGGCTGATGGGTGGGTGGGGCTGGCGGTG 32511
382  .....Th 382
32510  ACAAGATGTGCTCTTCCAGACCTGTGACCCATCTCATCCAGAC 32461
382  rGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLys 399
|||||
32460  AGGCGAGGTGCTGAAAGGGGGCTGTGCTTCAGTTGCATGACCTGAAC 32411
399  rGluAlaGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
|||||
32410  GGGAGGAGGAGGGGCTATCGCTGCTGGCTGTGCCCCACATACCC 32361
416  GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGly..... 429
|||||
32360  GGGCTGAACCGCACAGCTGTGTAACGTGCCATTTTGTGTAGGCCCT 32311
429  ..... 429
32310  CCTCTGGGGTAGAGACAGGTCAACCCCAAGTGGGTGTTTAACTC 32261
429  ..... 429

429  ..... 429
32260  TTTGACAAAAAGCACCTGTGCTGGGAGCTGTGGTGGAGGGGA 32211
430  .....Pr 430
32210  GGCAGGCTAGCTGGGCTTGCCACCTGACCTTCCTACCTGCCAAGGCC 32161
430  oProThrMetAlaPheLysGluArgLysValTrpValLysGluAsnMet 446
|||||
32160  CCCCTGGATGGCATTCAGAGAGAGAGAGAGGTGGGTGAAGAAGATATG 32111
447  ValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 463
|||||
32110  GTGTGAATCTGTCTGTGTGAAGGCTCAGGGACCCCGCCACATCTTC 32061
463  rTrpAsnValAsnGlyThr..... 469
32060  CTGGAGAGTCAAGGGCAGGGTGAAGCGGGGTCTCATCTGCCACATCTCC 32011
469  ..... 469
32010  TAACTCCTGCTGGGCTTGTGGCAGAGACAGATCTCCCATCCCC 31961
470  .....AlaSerGluGlnAspGlnAspThr 477
31960  TGAGCTGCTCTCTCTCTCTCTGCGCAGGCAAGTGAACAAAGCAAGATCC 31911
477  oGluArgValLeuSerThrLeuAsnValLeuValThrProGluLeuLeu 494
|||||
31910  ACAGCGAGTCTGTGACACCTGATGTCTCTGTGACCCCGGAGCTGTGG 31861
494  IuThrGlyValGluCysThrAlaSerAsnAspLeuGlyLysAsnThrSer 510
|||||
31860  AGACAGGTGTGAATGCAGGGCTCTCAAGACAGCTGGCAAAAACACAGC 31811
511  IleLeuPheLeuGluLeu..Val..... 517
|||||
31810  ATCTCTCTCTGAGAGCTGGGTGAGGGCTGCATCTCCCTCAGGGATGGCAG 31761
517  ..... 517
31760  GGGCAGGCCAGGACGAGGGGAGTAACCTTGGCTCTCTTAACCCCTCC 31711
518  .....AsnLeuThrThrLeuThrProAspSerAsnThr 529
|||||
31710  CTCTGCTCTCTAGTCAATTTAACCACTCTACACAGACTCTCAACACAC 31661
529  rThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaAsnSer 546
|||||
31660  CACTGGCTCAGCACTTCCACTGCCAGTCTCATACCAAGACCAACAGCA 31611
546  hrSerThr..... 548
31610  CCTCCACAGGTAAAGCCAGGCTGGCAAGAGAAAGGGCTGTGCCAGGCA 31561
548  ..... 548
31560  TCCCTTCTGCTCTGTCCCTCCCAAGAGCCCTGTGCCAAGAAAGTGAATA 31511
548  ..... 548
31510  GCAGCCCATCTTGTGGGCTGGAAGTGGGTGGGCAAGATGTGAGCA 31461
549  .....GluArg 550
31460  AGTGGCTGTGGGACGAGAGTGAAGAGAGTCTTTGTGGCGCAGAGAGA 31411
551  LysLeuProGluProGluSerArg..GlyValValIleValAlaValIle 567
|||||
31410  AAGCTGCCGAGAGCGAGAGCGGGGGCTGTGATCTGTGGGTGTGATGT 31361
567  aLysIleLeuValLeuAlaValLeuGlyAlaValLeuThrPheLeuTyr 583
```

```
|||||
31360 TGTGCAATCCCTGCTGCGGCTGCGGCTGCTGCTATTTCCCTAT 31311
584 LysylGlyLysLeuProCysArgArgSerGlyLysGlnIu.Ile.... 598
|||||
31310 AAGAAGGGCAAGCTGCTGCAAGGCGCTCAGGGAAGCAGAGATGTAAAC 31261
598 ..... 598
31260 CCAAGACGGGCCCCCTCCGGCCCCGCCCTTTTGCCACCAGCCTCTGCC 31211
598 ..... 598
31210 TGCCAGACGGCCCGCTGCCCCCTTGACAGCAGAGTGGGGCCCTGGGCTCAG 31161
598 ..... 598
31160 GGTGCTGCTGCTCCACTACCACTGCCAGCTCCATCAGCCACAGTGT 31111
598 ..... 598
31110 CTCATGACCTGCTCCCTGCTGGGCTGGCCCTGGAGCAGAGCCCTC 31061
598 ..... 598
31060 ATCCATTCACATCTTACACAGCGGGCCCTTCTCTCTGCTGCGCTGAG 31011
599 ..... ThreupProPseSa 604
31010 ATACCACTCTACCTCCCTCCCTCCCTCCAGCAGCAGTGGCCCCCTCTC 30961
604 rGlySerGluLeuValValGluValLysSerAspLysLeuProGlu 620
|||||
30960 GTAAAGCGCAACTGTAGTTAAGTTAAGTCAGATAAGCTCCAGAAAG 30911
621 MetGlyLeuLeuGlnLysSerSerGlyAspLysArgAlaProGlyAspG 637
|||||
30910 ATGGGCTCTCTGCAAGGCGCAGCAGCGGTGACAAAGAGGCTCGGAGACCA 30861
637 n 637
30860 G 30860
seq_name: gb_pr:HSWUC184
seq_documentation_block:
LOCUS HSMUC184 2015 bp DNA linear PRI 14-DEC-1994
DEFINITION H.sapiens MUC18 gene exons 6-13.
ACCESSION X68267
VERSION X68267.1 GI:602337
KEYWORDS adhesion molecule; cell surface molecule; immunoglobulin
superfamily; melanoma associated protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2015)
AUTHORS Sers,C
TITLE Direct Submission
JOURNAL Submitted (24-AUG-1992) C. Sers, Institut fuer Immunologie,
Universitaet Muenchen, Goethestr 31, 8000 Muenchen 2, FRG
REFERENCE 2 (bases 1 to 2015)
AUTHORS Sers,C, Kirsch,K, Rothbacher,U, Rietmuller,G, and Johnson,J.P.
TITLE Genomic organization of the melanoma-associated glycoprotein MUC18:
implications for the evolution of the immunoglobulin domains
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)
MEDLINE 93391384
FEATURES
source location/Qualifiers
1..2015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="leucocyte"
/clone_lib="HL 106d (Clonetechn)"
```

```
intron /clone="30B,17A,13"
<1..147
/number=5
exon 148..327
/number=6
intron 328..447
/number=6
exon 448..569
/number=7
intron 570..667
/number=7
exon 668..830
/number=8
intron 831..925
/number=8
exon 926..1044
/number=9
intron 1045..1146
/number=9
exon 1147..1288
/number=10
intron 1289..1443
/number=10
exon 1444..1565
/number=11
intron 1566..1676
/number=11
exon 1677..1818
/number=12
intron 1819..1915
/number=12
exon 1916..2011
/number=13
intron 2012..>2015
/number=13
BASE COUNT 424 a 592 c 606 g 393 t
ORIGIN
alignment_scores:
Quality: 1540.50 Length: 622
Ratio: 4.327 Gaps: 11
Percent Similarity: 57.235 Percent Identity: 56.431
alignment_block:
US-09-653-961-2 x HSMUC184 ..
Align seg 1/1 to: HSMUC184 from: 1 to: 2015
188 ValHisIleGlnSerSerGlnThrValGluSerSerGlyLeuTyrThrLe 204
|||||
150 GTCCACATTCAGTGTCCCAAGCTGTGAGTGAAGTGGTTGTAGACCTT 199
204 uGlnSerIleLeuLysAlaGlnLeuValLysGluAspLysAspAlaGlnP 221
|||||
200 GCAGAGATTCTGTGAAGGCACAGCTGTTAAAGAACAAAGATGCCACT 249
221 heTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnHisMetLysGlu 237
|||||
250 TTTTCTGTGAGCTCACCCTGCTCCAGTGGAGACCACTGAAGGAG 299
238 SerArgGluValThrValProValPheTyr..... 247
|||||
300 TCCAGGGAAAGTACACGTCCTGTTTCTGTGAGTACTAGCTGACTCT 349
247 ..... 247
350 GCTTACTGGGGCTGGGGCTGCTGGAACACTCTGCTGCCCTTATCC 399
247 ..... 247
400 AGCCCAAGGTGACTATGGGCCAAGAGCTACCCCTTTGCCCTAGCAGAC 449
```

248 ProThrGluLysValTrpLeuGluValGluProValGlyMetLeuLysGlu 264
 450 CCGACAGAAAAAGTGTGCTGGAAAGTGGAGCCCGTGGGAATGCTGAAGA 499
 264 uGlyAspArgValGluIleArgCysLeuAlaSpGlyAsnProProH 281
 500 AGGGGACCGCGTGAATCAGGTGTGCTATGGCAACCTCCACACCA 549
 281 AspSerIleSerLys..... 286
 550 ACTTCAGCAATCAGCAAGAGGTGTGGGGCTTGTGCTGGTGCAGGTGG 599
 286 286
 600 GGATGCAGACAGACAGCCAGGTGAGTGGCGGGAGGTGCATCTGCTCAC 649
 287Gln..AsnProSerThrArgGluAlaGluGluLys 298
 650 CTCGCCAATCTCTCCCTAGAACCCAGCAGCAGGAGGAGGAGAGAGAGAC 699
 298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
 700 AACCAAGACAGACAGGGGTCTGCTGGTGGAGCTGCCCGGAAGAACACA 749
 315 erGlyArgTyrGlyLysGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
 750 GTGGGCGCTATGATGATGAGCGCTGGAACTGGACACCATGATATGCTG 799
 332 LeuSerGluProGlnGluLeuValAsn..... 341
 800 CTGAGTACACACAGAACTACTGTGTAAGTGTGAGGGGCTGGGAGCCA 849
 341 341
 850 GGACAGGGGAGACAGGCTGGGGCAATAGGAAGTGGCCCACTGCTCCCT 899
 342Tyr..ValSerAspValArgValSerP 350
 900 GGGTGACATGATTGCTGTACCCCAAGATGTCTGACGTCCGAGTGAATC 949
 350 roAlaAlaProGluArgGlnGluLysSerLeuThrLeuThrCysGlu 366
 950 CCGCAGCCCTGAGAGACAGAGGAGGACACCTCACCTGCTGAGTGTGAG 999
 367 AlaGluSerSerGlnAspLeuGluPheGluTrpLeuArgGluLys 381
 1000 GCAGAGATGACCAAGACCTCGAGTTCAGTGGCTGAGAAAGAGTACC 1049
 381 381
 1050 CAGGGGGGTCTGGGGCAAGGGGCTGATGGGTGGGGCTGGCGGTGA 1099
 382Thr 382
 1100 CAAGATGTGCTCTTCCAGACCTGTGACCATCTCATCCAGACA 1149
 383 GlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysAr 399
 1150 GACCAAGTCTCGAAGAGGGGCTGTGCTGATGCAATGACACTGAACG 1199
 399 gGluAlaGlyGlyTyrArgCysValAlaSerValProSerIleProG 416
 1200 GGAGGAGAGGCGGCTATCGCTGGCTGCTGTGCGCCAGCAATATCCG 1249
 416 LysLeuAsnArgThrGlnLeuValAsnValAlaIle..Phe..... 428
 1250 GCCTGAACCGCACACAGCTGTGCAAGCTGCCATTTTGTGAGGCCCTC 1299
 428 428
 1300 CCTGTGGTAGAGACAGGTACACCCCAAGTGGGTGTTTAAAGCTCTT 1349
 428 428

1350 TGACAAAAAGCCACCTGCTGCGCTGGGGAGCTGTGTGCGGAGGGGAGG 1399
 429GlyProP 431
 1400 CAGGCTAGCTGGGCTTGCACCTGACCTCTCTACCTCCCAAGGGCCCC 1449
 431 roTPMetAlaPheLysGluArgLysValTrpValLysGluAsnMetVal 1447
 1450 CTGGATGGCATTCAGAGAGAGGAAGGTGTGGTGAAGAAGAAATATGCTG 1499
 448 LeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSerTr 464
 1500 TTGAATCTGCTGTGAGAGCGTCAGAGGACCCCGCGCCACCATCTCTG 1549
 464 PAsnValAsnGlyThr..... 469
 1550 GAACGTCAACGGCACGCTGAGCGGGGTCTCTATCTGCCACACTCTCTAA 1599
 469 469
 1600 TCCCTGCTGGCGCTTGTGGCCAGGACAGATCTCCCATCCCTGA 1649
 470AlaSerGluGlnAspGlnAspPr 477
 1650 GCTGCTCTCTCTCTCTCTCTCTCTCGCAGGCAAGTGAACAGACAGATCC 1699
 477 oGlnArgValLeuSerThrLeuAsnValLeuValThrProGluLeuG 494
 1700 ACAGCAGTCTGAGCAGCACCCTGAATGCTCTGTGACCCGGAGCTGTGG 1749
 494 LuThrGlyValLysCysThrAlaSerAsnAspLeuGlyLysAsnThrSer 510
 1750 AGCAGGTGTGAATGACAGGCGCTCCAAAGCCTGGGCAAAAAACACAGC 1799
 511 IleLeuPheGluGluLeu..Val..... 517
 1800 ATCTCTCTCTGAGCTGGGTGAGGCTGCATCTCTGCAGGGCATGGGCA 1849
 517 517
 1850 GGGGCTCAGGGCCAGGCGGAGTAACTGCGCTTCTCTACACCCC 1899
 518AsnLeuThrThrLeuThrProAspSerAsnTh 528
 1900 TCCCTGCTCTCTCTAGTCAATTAACCACTTACACAGACTCCACAGC 1949
 528 rThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaAsn 545
 1950 AACCACTGGCTCAGCACTTCCAGTCCAGTCTCTCATACAGAGCCACA 1999
 545 erThrSerThr 548
 2000 GCACCTCCACA 2010

seq name: gb.ro:MMUC18GP

seq documentation block: 6102 bp DNA linear ROD 05-OCT-1993

LOCUS MMUC18GP

DEFINITION M.musculus gene for MUC18 glycoprotein (partial).

ACCESSION X74628

VERSION X74628.1 GI:406585

KEYWORDS immunoglobulin gene superfamily; melanoma associated protein; MUC18 glycoprotein.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 6102)

1 Rothbacher, U. Direct Submission

Submitted (17-AUG-1993) U. Rothbacher, Institute for Immunology,

University of Muenich, Goethestr 31, 80336 Muenchen 2, FRG

REFERENCE 2 (bases 1 to 6102)
AUTHORS Rothbacher,U., Sers,C., Riethmuller,G. and Johnson,J.P.
TITLE Characterization of the human melanoma metastasis associated
molecule MUC18 in the mouse. Unique gene structure and expression
in murine melanomas

JOURNAL Unpublished
FEATURES Location/Qualifiers

1..6102

/organism="Mus musculus"

/strain="BALB/c"

/db_xref="taxon:10090"

/dev_stage="adult"

/tissue_type="liver"

/clone="lambda 16, lambda 17, lambda 35"

/clone_11b="EMBL-3 SP6/T7 (Clontech No.ML1030)"

1..901

/note="immunoglobulin like (V-set) domain II"

647..652

774..953

/label=ex5

1097..1270

/label=ex6

1147..1452

/note="immunoglobulin like (C2-set) domain III"

1339..1501

/label=ex7

1592..1710

/label=ex8

1636..1891

/note="immunoglobulin like (C2-set) domain IV"

1796..1937

/label=ex9

1904..1906

/note="glycosylation (N-linked)"

2072..2193

/label=ex10

2116..2411

/note="immunoglobulin like (C2-set) domain V"

2131..2133

/note="glycosylation (N-linked)"

2185..2187

/note="glycosylation (N-linked)"

2301..2442

/label=ex11

2415..2417

/note="glycosylation (N-linked)"

2547..2632

/label=ex12

2617..2619

/note="glycosylation (N-linked)"

2784..2931

/label=ex13

2811..2872

/note="transmembrane domain"

3218..3335

/label=ex14

3218..3227

/note="phosphorylation site (PKC)"

3231..3238

/note="phosphorylation site (PKC)"

3234..3262

/note="phosphorylation site (PKC)"

3303..3314

/note="phosphorylation site (CKII)"

3727..3756

/label=ex15

4647..4652

5010..5015

5772..5777

6027..6032

1577 c 1520 g 1545 t

alignment_scores:

Ratio: 1443.00

Percent Similarity: 48.775

alignment_block:

US-09-653-961-2 x MMUC18GP ..

Align seg 1/1 to: MMUC18GP from: 1 to: 6102

188 ValHisIrlGlnSerSerGlnThrValGlnSerSerGlyLeuYrThrLe 204

776 GTCATATCCAGTCATCACATGTCGAGTCCTCAGTCCTGTCACCTT 825

204 uGlnSerIleLeuYsAlaGlnLeuValYsGlnYsGlnYsAlaGln 221

826 GAGAGCTGTCTGAGTCAGCGCTGATTAAGGAGACAAAGATGCCAGT 875

221 heYrCysGlnLeuAsnYrArgLeuProSerGlyAsnHisMetYsGlu 237

876 TTTACTGTAACTCAGTACCGCTACCCAGTGGGAGACATGAGGAA 925

238 SerArgGluValThrValProValPhe..... 246

926 TCTAAGAGAGTCATCTGCTCTTTTCTGTAGTGTGACTTCTAGCGC 975

246

976 CTCTGATTTTGTGAGGGGGGAGATGGGGTGGGCGCTGTGGGAGAC 1025

246

1026 TCCTCTGCTGCTCTGTGTCAAAATGGGGCCACCTTGGAGCTGAGAG 1075

247

1076 CCTTCTCTTGGCCAAACAGACCTGAGAAAGTGTGGTGGAGGAT 1125

257 GluProValGlyMetLeuYsGlnGlyAspArgValGluIleArgCysLe 273

1126 GAGCGTGTGGGCTGTGAAGGAGGATCATGTGCAATTCAGTGCTC 1175

273 uAlaAspGlyAsnProProHisPheSerIleSerIys..... 286

1176 GACAGATGCAACCCCTCACCCCTTCACTATCAACAGAAAGTACAGA 1225

286

1226 GAGAGCGAGCGCTCCATGCTGGGTGACGAGGTGGCGTAAAGGTGTGGG 1275

286

1276 CAGAGGTGGGAGACGTAAGTGAAGGCATCAGCTCCTCACCCCTTT 1325

287

1326 GTGCTCCCCAGAGCCCGACCTGAGGAGATGGAAGAGAGAGACCG 1375

300 sNaSPAsnGlyValLeuValLeuGluProAlaArgYsGluHisSerGly 316

1376 ATGAAATGGGCTGCTGCTGAGCCCTGCGCAAAAGACCATAGCGGG 1425

317 ArgYrGlnCysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeu 333

1426 CTACACAGTGTGAGAGTGTGAGCTGGAATACATACACTGTCAAG 1475

333 rGluProGlnGluLeuValAsnYr..... 342

1476 TGACCCCTGTGAGCTGTGTGAACTGTGAGACCTGGAGTGTGACAG 1525

342

342

1526 CTCACCTCTGGGCAAGAGGCAACTGGCTCCTCCTGTCCAGGCTGACATTTT 1575
1526 CTCACCTCTGGGCAAGAGGCAACTGGCTCCTCCTGTCCAGGCTGACATTTT 1575
343ValSerAspValArgValSerProAlaIaIaPr 353
1576 TTTTGCTATGGCTCCAGATGTGTCTGATGATTTCAAGTCAATCAACCTGCCCC 1625
353 oGUATyGlnGluGlySerSerLeuThrLeuThrCysGlnAlaGluIleuS 370
1626 TGAAGTCCAGAGAGGTAGAGACCCACAGCTGACCTGGCGAGGCAAGAGTA 1675
370 ergIlaSpleuGluPhgGlnIrrPheIuArgIlu.Glu..... 381
1676 ACCAGACCTTGAAGTTTGAGTGGCTGAGACAGACAGTAACCAACACTT 1725
381 381
1726 GAGACTGGGAGAGGGGAGCAATGAGACACTTGCCTCTTTTCAGACCCCT 1775
382 ThrGlyGlnAlaLeuGluArgIlyProVal 391
1776 GACCACCTCTCACCCCTAGACAGGCCAGCTGTGTGGAAAGGTCGCTC 1825
392 LeuGlnLeuHisAspLeuLysArgIluIaGlyIylYrArgCysVa 408
1826 CTCGACGTAACAAACAGTGAGAGACGAGAACAGGAGGAGCGATATCTGCAT 1875
408 lAlaSerValProSerIleProGlyLeuAsnArgThrGlnLeuValAsnV 425
1876 GGCATCTGTCCCCAGAGATTCCTGGCTTGATATCTGATCCACAGCTGTGACGG 1925
425 aIaIaIlePheGlyPro..... 430
1926 TGGGATTTTGTGTGAGGCCCTTTGTACAGTACACAGTCCACCTGAGTT 1975
430 430
1976 GTGTGGTTTGAAGCTGTGGCCAGCTTGTACTGTGGAGCTAGCGGGGA 2025
431P 431
2026 GCAGGCTATTGGTATTGGGGGGGTAAACCTTCCTGCCCAAGGTC 2075
431 ro..TrpMetAlaPheLysGluArgLysValTrpValLysGluAsnMetY 447
2076 CCATGATGATGGCATTAAAGAGAGGAGAGGTGTGGGTCAAGAAATGACG 2125
447 alLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSer 463
2126 TGTGTATCTGTCTTGTGAGCTTCAAGACATCTCAAGCCACATCTCC 2175
464 TrpAsnValAsnGly..... 468
2176 TGGAAATGATCAATGTTGCGTTCGGTAGACCGGCTTGCTCATCTCCCAT 2225
468 468
2226 TCTCTTAATCCCTCGAGCTTGGAGACCAAGCCCAAGATCGTTCGGCC 2275
469Thr..AlaSerGluGlnAspGlnAspProG 478
2276 CTCGTGAGAGGCCCATTTACTTCTAGGCAACGTAGATGGAACCAATCCAC 2325
478 lnaArgValLeuSerThrLeuAsnValLeuValThrProGluLeuGln 494
2326 AGACAGTAGTAGACCTTGATGTCTTGTGACGCCAGACACTCTGGAG 2375
495 ThrGlyValGluCysThrAlaSerAsnAspLeuLysAsnThrSerI 511
2376 ACAAGTGAAGGTGTACAGCTCCAACTCCCTGGGCTCAACACCAACAC 2425
511 e..... 511
2426 CATTTGTTCTGAAGCTGGGTGAGACGGACCTCCCTGCAGAGGCTGGACAGG 2475

512	leuphe	513
2476	GCATGCCAGGACGACAGGCACTTAACTTGCTGCTACTAAACCCCAACTCTT		2525
514	leuGLu.leuValaInleuThrThrleuThrProAspSerAsnThrThT		530
2526	CTGGCTCTGACTACTTAAACCACCTCATRACCTGACTCCAGCCAAAGCA		2575
530	hrcIyLeuSerThrSerThrIaSerProHisThrIraIaInSerThr		546
2576	CTGGCTCGAGCACTCTCCACAGTCAGTCCTCCACACGAGCAACACAGCAC		2625
547	.Ser.....		547
2626	TCCACAGGTAGCGGGCCCTGGCGAGTGAACCAAGTCTCCCTTT		2675
547		547
2676	TGAGAAAGTTAAAGGCCACTCTGTCCCTGGGCTGGAGAGGGGGGCA		2725
547		547
2726	GCAACAGCAATTTGGCAGCCAGGACGAGGAATGACTCAAATCTTTG		2775
548	...ThrGluArgLysLeuProGluIupProGluSerArgLysValValIle		563
2776	TTTGACAGAGAAAAACCTCCACAGCCAGAGGCAAAAGTGTGTCATCG		2825
563	AlAlaValIleValCysIleLeuValIleuAlaValIleuValIleu		579
2826	TGGCTGTGATAGTGTGTACTCTGGTCTGCTGCTGCTGGGCTGCTCTC		2875
580	TyrPheLeuIyrIysLysGlyLysLeuProCysAlaArgSerGlyLysG		596
2876	TATTCTTCTCAAGAGAGGCAAGCTGCCATGTGAGACCTGGGAAAAA		2925
596	nglu.Ile.....		598
2926	GGAGGTGAAGCCAGGCTTATGCGCTCTCCACCCCAACCCCAATTTTG		2975
598		598
2976	GCGTCCAGCTGCTGCTGCTTATTAAGGGCCCTGCTACCCAGTGGCCATT		3025
598		598
3026	ACCTGGAGAGCAAAACTGAGACCTGTTTATTATGTCACACTCAAACTG		3075
598		598
3076	CCACCAACCAACCACTCTCCCAATTCCTTCTGAGCGGCCCA		3125
598		598
3126	AAAGCAGCAGATATCCATTCTTAAAAAGTCCCCCCCCCTCCAGGCTTC		3175
599		599
3176	CCACGCCCCGAGTTGGCCAACTCTCACTCCCTCTCCAGACAGCTTC		3225
601	roProSerArgLysSerGluLeuValValGluValLysSerAspLysLeu		617
3226	CCCCGCTGTGAAGGTAAATTTGTAGTTGAAGTTAAGTCAAGTAAGTC		3275
618	ProGluIuMetGlyLeuLeuGlnGlySerSerGlyAspLysArgAlaI		634
3276	CCAGAAAGATAGTCTCTCTTCAGGAGCAACGCTGACAAAGAGGCTCC		3325
634	oGlyAspGln 637		637
3326	AGGAGCCAG 3335		3335

seq_name: gb_hcg:AP002888

seq_documentation_block:

LOCUS AP002888 156869 bp DNA linear HTG 18-OCT-2000

DEFINITION Homo sapiens chromosome 11 clone RP11-122H4 map 11q, WORKING DRAFT

SEQUENCE, 25 unordereded pieces.

ACCESSION AP002888.1 GI:10880444

VERSION HTG; HTGS-PHASE1; HTGS-DRAFT.

KEYWORDS Homo sapiens DNA, clone:RP11-122H4.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 156869)

Autor: Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,

Funayama, A., Yada, T., Totoki, Y., Matanabe, H. and Sakaki, Y.

Homosapiens 156,869 genomic DNA of 11q

2 (bases 1 to 156869)

Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,

Funayama, A., Yada, T., Totoki, Y., Matanabe, H. and Sakaki, Y.

Submitted (16-OCT-2000) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan

(E-mail: hattori@igsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,

TEL:81-42-778-9923, Fax:81-42-778-9924)

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@igsc.riken.go.jp

----- Project Information

Center project name: RP11-122H4

----- Summary Statistics

Sequencing vector: PCR products: 100% of reads

Assembly program: Phrap, version 0.990329

Consensus quality: 137843 bases at least Q40

Consensus quality: 151403 bases at least Q30

Insert size: 154469, sum-of-contigs

Quality coverage: 4.53x in Q20 bases, sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of

25 contigs. The true order of the pieces is not known and their

order in this sequence record is arbitrary. Gaps between the

contigs are represented as runs 'N', but the exact sizes of the gaps

are unknown. This record will be updated with the finished sequence

as soon as it is available and the accession number will be

preserved

```

1 16262 contig of 16262 bp in length
1633 32533 contig of 16171 bp in length
32634 47033 contig of 14400 bp in length
47134 60266 contig of 13133 bp in length
60367 70550 contig of 10184 bp in length
70651 83136 contig of 12486 bp in length
83237 90994 contig of 7758 bp in length
91095 99344 contig of 8250 bp in length
99445 106022 contig of 6578 bp in length
106123 112604 contig of 6482 bp in length
112705 123495 contig of 5570 bp in length
118375 135477 contig of 5121 bp in length
123596 130199 contig of 6604 bp in length
130300 132629 contig of 2330 bp in length
132730 135477 contig of 2748 bp in length
135578 139010 contig of 3433 bp in length
139111 141568 contig of 2458 bp in length
141669 142131 contig of 463 bp in length
142232 144866 contig of 2635 bp in length
144967 148144 contig of 3448 bp in length
148145 150504 contig of 1990 bp in length
150505 150604 contig of 100 bp in length
150605 151939 contig of 1355 bp in length
151960 152059 contig of 100 bp in length
152060 153455 contig of 1396 bp in length
153456 153555 contig of 100 bp in length
153556 155251 contig of 1696 bp in length
155252 155351 contig of 100 bp in length
155352 155869 contig of 1518 bp in length.

```

FEATURES

source

1. 156869

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11q"

/clone="RP11-122H4"

1. 16262

/note="assembly-fragment"

16363. 32533

/note="assembly-fragment"

32634. 47033

```
misc_feature /note="assembly_fragment"
47134..60266
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
60367..70550
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
70651..83136
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
83237..90994
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
91095..99344
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
99445..106022
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
106123..112604
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
112705..118274
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
118375..123495
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
123596..130199
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
130300..132629
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
132730..135477
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
135578..139010
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
139111..141568
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
141669..142131
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
142232..144866
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
144967..148414
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
148515..150504
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
150605..151959
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
152060..153455
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
153556..155251
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
155352..156869
/note="assembly_fragment"
BASE COUNT 42936 a 35209 c 34104 g 42199 t 2421 others
ORIGIN

alignment_scores:
Quality: 1202.00 Length: 1059
Ratio: 3.588 Gaps: 12
Percent Similarity: 31.634 Percent Identity: 31.539

alignment_block:
US-09-653-961-2 x AP002888 ..
Align seg 1/1 to: AP002888 from: 1 to: 156869

19 ProArvYal.....AlaclyValProGlyGluAlacGlnProAlaPr 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145244 CCGCGCTCCCTCCCGAGGTGCGCGAGAGCTGAGCAGCCTCGGCC 145293
33 oGluLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyL 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145294 TGAGCTGTGAGGTGAGAGTGGGAGCAGCAGCCCTGTGAAGTGGGCC 145343
50 euSerGlnSerGlnGlyAsnLeuSerHisValAspTrp..PhaSer.... 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145344 TCTCCAGTCCCAAGGCAACCTCAGCAGATGCTGACTGTTCTGTGAGT 145393
64 ..... 64
```

```
145394 GCTTGGGCTCTCTGCAAGAGAGCCCGGAGGAGAGACTCTGCCCTCTTG 145443
65 .....ValHisLy 67
145444 TAGCCAGAGACCTTCTCTCTCTGTGTCTCTCTCCATTAGTCCACAA 145493
67 sGluYsArqThrLeuIlePheArqValArgGlnGlyGlnGlyInsrg 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145494 GGAGAAAGCGAGCTCATCTTCCGTGTGGCCAGAGGCCAGAGGCCAGAGCG 145543
84 lUpProGlyLysArgLysGlnArgLeuSerLeuGlnAspArgLysAlaThr 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145544 AACCTGGGAGTACAGAGCGGCTCAGCCTCCAGAGCAGAGAGGCTCTCT 145593
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgLlePheLeuCyseL 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145594 CTGGCCCTGACTCAAGTCAATCCCCCAAGACGAGCGATCTCTGTGCCA 145643
117 nGlyYsArqProArqSerGlnGlyArgLleGlnLeuArgValYr. 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145644 GGNCAAGCGCCCTCGGTCCAGAGATACCGATCCAGTCCCGCTTACA 145693
133 ..... 133
145694 GTGCGTAGCCCTGTGTGTGCCAGAGGATGAGGCTGTGGACGTGG 145743
134 .....LysAlaProGluLupr 139
145744 CCAGATGTCCCTCATTTCTATTCTGTGTAGAGCT..CGGAGAGAGCC 145792
139 oAsnIleGlnValAsnProLeuGlyIleProValAsnSerLysGluProG 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145793 AAACATCCAGGTCAACCCCTGGGATCCCTGTGAACAGTAGAGGCTG 145842
156 lUglu..... 157
145843 AGGAGGTGAGATGTGTGCGAGGTGGGCACTGTGTCCAGCCGCT 145892
157 ..... 157
145893 GAGATGGGCTTGAAAGGGGCTCGAGAGCAGGAGACTCACCCCTGTGC 145942
158 .....ValAlaThrCysValGlyArgAsnGlyTyrProIleProGlnV 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145943 TCTTAGGTGCTGCTACCTGTGTAGAGGAGAAACGGGTACCCCATTTCTCAAG 145992
172 aLleTrpThrLysAsnGlyArgProLeuLysGlnLysAsnArg... 187
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145993 TCATCTGTATCAAGAAATGGCCGGCTCTGAAGAGGAGAAACGCTGAG 146042
187 ..... 187
146043 TAGCCCTTCTCCAAAAGCCACCCCTTGGGCCAGGTGTGTAGCTCACAC 146092
187 ..... 187
146093 CTGTAAATCTGGTCTTTGGAGGCCAAGCAGAGAGATCTGTGAGGTG 146142
187 ..... 187
146143 AGGAGTTCAGACCACTTGGACAACAATAGGAGCTGTCTACAGA 146192
187 ..... 187
146193 AAATATAAAATTTAGCAGAGCATGTGGTGTGTGCTGTGTCCAGATA 146242
187 ..... 187
146243 CTGGGAGGCTGAGGAGAGAGATCATATGAGCCAGAAATGAGGCTG 146292
187 ..... 187
146293 CAGTAGCTAGAGATTGCACACACACTCCGGGCTGTGTGACAGAGCAG 146342
```

187 187
146343 ACCCATCTCTTAAAAAGAGCTGTCCCTATGTAAGTACACAGAACTGCC 146392
187 187
146393 TTATATTTTCATTAGCTTTTAACTTAAAAAATTATACAAATGTACTT 146442
187 187
146443 GAAATATTTTCATAGAAATTAAACGACAAAAAATCTAAGATTCTT 146492
187 187
146493 CTTCCTGAGACCCCCCTCCAGAAATATCTGTGTAAGGCTGTG 146542
187 187
146543 TCTTCAAGCCATATTCTATGTATTTGCTCATATATAGAAATATTCTAG 146592
187 187
146593 AATGATATAGGCTTCTGTGTATTATCTAAATCAGTCATTCYTAACA 146642
187 187
146643 GGGGTGATTTTGTACCCCTCTCTCTAGAGATCTGGAATGTGGA 146692
187 187
146693 GATATTTTGTGTGCACACATAGAGGGGTGCTACTGCCATCTAAGAG 146742
187 187
146743 TAGAGAGACCAAGATGTGCTAATCCTATAGGACAGACAGCCCC 146792
187 187
146793 CACATATTAAGATCAACGTGTGCTTAAACATCAGTAGTGTGCTGGGC 146842
187 187
146843 TCACGCCCTGTAATCCACACTTTTGGAGCCAAAGTGGGGATCACCT 146892
187 187
146893 GAGGTGGAGTACAAAGACAGCCTGTCCACACAGAGAAACCCATCTC 146942
187 187
146943 TACTGAAATACAAAAGTAGCCGGCGGTGGCGCATGTCTGTAATCCC 146992
187 187
146993 AGCTACTGACGAGCTGAGAGAGAAATCATTGAAAGCCGAGAGGAG 147042
187 187
147043 TGGAGGTTTCGGTGAAGCCAGATTTACCCCTGCACCTCAGCGTGGCCA 147092
187 187
147093 CAAGAGTGAACCTTTTAAAAAATAAAAAAAAAAATATCAGTAGTG 147142
187 187
147143 CTGAGAAACCTGTGTAAAGTGTGTGTATACATTTGTTAGACAT 147192
187 187
147193 TTCTTTATACAATGTTCTGGGTCAAGTCTATTAGATCAACTGATGTT 147242

187 187
147243 TTGCTTACTGCCAAGTTTTCATCTACTACGATAGCAGTAGTCGAGTTCA 147292
187 187
147293 CCATTCCCATTTAGTGGACATCTAGACGGCTGCTGTTTATCATTTGC 147342
187 187
147343 AGCATCTTTGCACACATCTTGATATAGCAGACGAAATGTTTTT 147392
187 187
147393 CTAGGTTGCACATGACGATAAAGTCTGGGTGTAAGGTTTCCAGCT 147442
187 187
147443 TGCATTTGTAAGCTGGCTTTCACAGGGGACAGGGGCTATTAGATGTC 147492
187 187
147493 CCTGCCAACCCAGTGGACAAACCTAGGGTGGGCTGAGGTGGGCTG 147542
188 ValHisIleGlnse 192
147543 AGGTGAGTCTTCTCCCTTCTCCCTGCCAGGGGCTCACATTCAGTC 147592
192 rserGlnThrValIgluserSerGlyLeuYrThrLeuGlnSerIleLeuL 209
147593 GTCCAGACTGTGAGTGCAGTGGTTTGACACCTTGCAGAGATATCTGA 147642
209 ysAlaGlnLeuValIySgluAspIyAspAlaGlnPheYrCySgluLeu 225
147643 AGGCACAGCTGGTTAAAGAACAAATCCAGATTACTGTCGACGTC 147692
226 AsnTyRArgLeuProSerGlyAsnHisMetIySgluSerAArgIyValTh 242
147693 AACTACGGGTGCCAGTGGAACACATGAAAGAGTCCAGGAAAGTAC 147742
242 rValProValPheTyR..... 247
147743 CGTCCGTGTTTCTGTGAGTACTAACCTGACTCTCTACTGGGCGTG 147792
247 247
147793 GGGCTGTGGAAACTCTGCTGCTCCCTTATCCAGCCAGGGTTGAC 147842
248 ProThrGluIySva 252
147843 TCATGGGCCAAGAGCTACCCCTTGTCCCTATGCAGACCCGAGAAAAGT 147892
252 LTrPLeuGluValIgluProValIgluMetLeuIySgluIyAspArgValG 269
147893 GTGGCTGGAAGTGGAGCCCGTGGGAATCTGAAGGAAGGGGACCCGCTG 147942
269 IuIleArgCySleuAlaAspIyAspProProHisPheSerIleSer 285
147943 AATCAGGTGTTTGGCTATGCAACCTTCACACACTTCAGCATCAGC 147992
286 Lys..... 286
147993 AAGCAGGTGGGGGCTTGTGCTGGGTGCAGGTGGGATGCACAGAG 148042
287 G 287
148043 ACCCAGGTGAGTGGCGGAGGTGCACCTGTCACCTCCCAATCTCC 148092
287 In..AsnProSerThrArgGluAlaIgluIySgluIyThrIyAsnAspArg 303
148093 CTAGAACCCCGACACAGGAGGAGGAGAAAGAACCAACCAACGCAACG 148142
303 IyValIleuValIleuGluProAlaIySgluHisSerClyArgTyGlu 319


```
|||||
148143 GGCTCTGCTGCTGAGGCTGCGGGAAGAACACAGTGGCGCTATGAA 148192
320 CysGlnGlyLeuAspLeuAspThrMetLysSerLeuSerGluProG1 336
|||||
148193 TGTACAGGCGCTGAGCTTGACACCATGATATCGCTGAGTGAAACACA 148242
336 nGluLeuLeuValasn..... 341
|||||
148243 GGAAGTACTGTTGACACTGTGAGGGGCTGGGAGCCAGGACAGGGGACCA 148292
341 ..... 341
148293 GGCTGGGGCAATAGGGAACCTGCCCCACCTGCGGTGACATGATTG 148342
342 .....Tyr..ValSerAspValArgValSerProLala1a.ProG1 354
|||||
148343 CTGTACCCCCCAGATGTGTCTGACGTCCGAGTGAAGTCCCGCAGCACACTGA 148392
354 uArgGlnGlySerSerLeu 361
|||||
148393 GAGACAGGAAGGACAGCGCTC 148414
```


OM of: US-09-653-961-2 to: N_Geneseq_032802.* out-format : pfs

Date: Jun 28, 2002 12:13 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frim+ .p2n.model -DEV=xlp  
-O=/cgn2_1/USPTO.spool/US09653961/rnmat.26062002.083939.764/app-query.fasta.1.712  
-DB=N_Geneseq_032802 -GFM=fastap -SUPFLX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsnum62  
-TRANS=human40.cdl -LIST=45 -DOCLALIGN=200 -THR SCORE=Pct  
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09653961.ecgn1.1.684 -NCPU=6 -ICPU=3 -LONGLOC  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-653-961-2  
Query length: 646  
Database: N_Geneseq_032802.*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 270.080000
```

Score_list:

Sequence	Strd Orig	ZScore	E_Score	len	Documentation	
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20930			3363.00	4329.00	8.8e-233	19
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20931			3327.00	4282.46	3.5e-230	19
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH02921			3327.00	4277.51	6.5e-220	19
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20932			3316.00	4266.24	2.1e-229	19
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20933			888.00	1144.93	2.0e-55	52
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20934			698.00	901.60	7.2e-42	44
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20935			682.00	880.51	1.1e-40	44
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20936			655.00	846.57	8.3e-39	40
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH83083			647.50	819.78	2.6e-37	40
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH83116			647.50	819.78	2.6e-37	40
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH83116			647.50	819.78	2.6e-37	40
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20937			606.00	765.36	2.9e-37	37
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20938			501.00	650.51	6.9e-28	29
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20939			499.00	622.57	2.5e-26	4
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAZ20940			498.50	628.26	1.2e-26	21
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAZ20941			497.00	624.87	1.9e-26	25
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAZ20942			494.50	624.66	1.9e-26	25
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAZ20943			466.00	588.05	2.1e-24	18
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20944			414.50	521.95	1.0e-20	17
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20945			292.00	364.80	5.7e-12	15
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ20946			292.00	364.80	5.7e-12	15
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ20947			280.50	343.97	8.2e-11	7
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ20948			280.50	343.97	8.2e-11	7
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ20949			269.50	334.96	2.6e-10	1
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ20950			269.50	334.96	2.6e-10	1
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAZ20951			255.00	311.82	3.1e-09	2
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ20952			252.00	303.76	1.4e-08	41
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20953			252.00	303.76	1.4e-08	41
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20954			251.50	291.28	7.1e-08	41
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20955			247.50	302.86	1.6e-08	2
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAZ20956			244.50	301.31	2.0e-08	19
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:AAZ20957			242.00	295.92	3.9e-08	24
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20958			242.00	295.92	3.9e-08	24
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH33526			241.50	297.43	3.2e-08	1
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:AAZ20959			241.50	295.62	4.1e-08	23
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20960			241.50	295.62	4.1e-08	23
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAO82807			241.00	296.05	3.8e-08	20

seq_name	/SID5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAZ20930	241.00	295.51	4.1e-08
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ20931		241.00	294.81	4.5e-08
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAZ20932		241.00	292.87	5.8e-08
/SID5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAZ20933		240.50	292.54	6.0e-08
/SID5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:AAZ20934		240.00	291.87	6.6e-08

seq_documentation_block:

ID	AAZ20930 standard; cDNA; 1950 BP.
AC	AAZ20930;
DT	01-DEC-1999 (first entry)
XX	Human MUC18 cDNA sequence.
DE	prostate cancer; melanoma; cell adhesion; glycoprotein;
KW	metastasis; treatment; detection; diagnostic test; ds.
XX	Homo sapiens.
OS	Homo sapiens.
PH	Key
FT	CDS
FT	1..1941
FT	/*tag= a
FT	/product= HumanMUC18
FT	/note= "Human MUC18"
XX	W09945392-A1.
XX	10-SEP-1999.
PD	02-MAR-1999; 99WO-US04850.
PF	03-MAR-1998; 98US-0076664.
PR	(UYEM-) UNTIV EMORY.
XX	MU G;
XX	WPI: 1999-540899/45.
XX	P-PSDB; AAZ2404.
PT	Detection of metastatic prostate cancer, by detection of MUC18
PT	expression in prostate cancer cells -
XX	Claim 18; Page 7; 80pp; English.
XX	This is the nucleotide sequence of the Human MUC18 cell adhesion
CC	glycoprotein, which is expressed on the surface of melanoma cells,
CC	and can be used as a marker for prostate cancer.
CC	This MUC18 (humMUC18) cDNA sequence is different from the humMUC18 cDNA
CC	given in Genebank accession number AAN28882, AAZ20931. The deduced amino
CC	acid residues from this cDNA are very different from that published by
CC	Johnson's group in 1989.
CC	The presence of this glycoprotein has been correlated with the ability
CC	of melanomas to metastasize. MUC18 is also associated with normal
CC	vascular tissue, and on the smooth muscle of venules, and it expresses
CC	sporadically on capillary epithelium.
CC	The method can be used as a diagnostic test for prostate cancer which
CC	has a relatively high potential for metastasis or which has
CC	metastasized. This allows the physician to choose appropriate surgical,
CC	radiation, or chemotherapeutic treatment regimens. In addition, antibody
CC	specific to MUC18 can be used to prevent metastasis of Prostate Cancer
CC	cells.
XX	Sequence 1950 BP; 454 A; 552 C; 596 G; 348 T; 0 other;

alignment_scores:

alignment_scores:	Quality: 3363.00	Length: 646
Ratio: 5.206 <td></td> <td>Gaps: 0</td>		Gaps: 0
Percent Similarity: 100.000		Percent Identity: 100.000

alignment block:
US-09-653-961-2 x AA220930 ..

Align seg 1/1 to: AA220930 from: 1 to: 1950

```
1 MetGlyLeuProAlaGlyLeuValCysAlaIleuLeuLeuAlaIaCysCysCys 17
1 ATGGGGCTTCCAGAGCTGGTCTGCGCTTCTGCTGCGCGCTGCTGCTG 50
17 SCysProAlaGlyAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
51 CTGTCTCTGCGCTGCGGGGTGTGCGCGAGAGCTGAGAGCTGCGGCTG 100
34 IuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
101 AGCTGGTGGAGGTGGAGAGTGGGACAGCACAGCCCTTGTGAAGTGGCGCTC 150
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLys 67
151 TCCAGTCCCAAGAGCAACCTCAGCCATCTCCAGTGGTTTCTGTCCACAA 200
67 SGIuLysArgThrLeuIlePheArgValArgGlnGlyGlnGlyGlnSerG 84
201 GGAGAAGCGGAGCGCTCATCTTCCGTGCGCAGAGCGCAGGGCCAGAGCG 250
84 IuProGlyGluTyrGluGlnArgLeuSerLeuGlnAspArgGlyValAlaThr 100
251 AACCTGGGAGTACGACGCGGCTCAGCCTTCACAGACAGAGGGGCTACT 300
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCysGly 117
301 CTGGGCGCTGACTCAAGTCAACCCGCCAAGACGAGCGCATCTTCTGTGCCA 350
117 nGlyLysArgProAlaGserGlnGlyTyrArgIleGlnLeuArgValTyrL 134
351 GGGCAAGCGCCCTCGTCCAGAGTACCGCATCCAGCTCGCGCTGTACA 400
134 yAlaIleProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
401 AAGCTCCGGAGAGGCCAAGATCCAGGTCAACCCCTGGGCATCCCTGTG 450
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTyr 167
451 AACAGTAAGAGCCTGAGAGAGTCCGTACTGTGTAGGAGGAAACGGGTA 500
167 rProIleProGlnValIleTrpTyrLysAsnGlyArgProLeuLysGluG 184
501 CCCCATTTCTCAAGTCAATCTGGTACAAAGATGGCGGCTCTGAAAGGAGG 550
184 IuLysAsnArgValHisIleGlnSerSerGlnThrValGluIleSerSerGly 200
551 AGAAGAACCGGGTCCCATCTAGTGTGCTCCACAGCTGTGAGTGCAGTGT 600
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAspLys 217
601 TTGTACACCTTGTGAGATTTCTGAAGGCACAGCTGGTTAAAGAGACAA 650
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnH 234
651 AGATGCCCAAGTTTACTGTGAGCTCAACTACCGGCTGCCAGTGGGAACC 700
234 IseMetLysGluSerArgGluValThrValProValPheTyrProThrGlu 250
701 ACATGAAGGAGTCCAGGAAAGTCAACGCTCCGTGTTTCTACCCGACAGAA 750
251 LysValTTrpLeuGluValGluProValIleGlyMetLeuLysGluLysPsr 267
751 AAAGTGTGGCTGGAAAGTGGAGCCGTGGGAATGCTGAAGGAAGGGAGCCG 800
267 yValGluIleArgCysLeuAlaAspGlyAsnProPropOHIsPheSerI 284
801 CGTGAATATCAGGTGTTTGGCTGATGGCAACCTCCACCACTTACAGA 850
```

```
284 IeSerLysGlnAsnProSerThrArgGluAlaGluGluGlnThrThrAsn 300
851 TCAGAAAGCAGAACCCAGCACACAGGAGGACAGAGAAAGAACACACAC 900
301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGlyAr 317
901 GACACAGGGGTCTGTGTGTGAGAGCTGCCCGGAAGAACACAGTGGGCG 950
317 gTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSerG 334
951 CTATGAATGTCAAGGCGCTGAGCTTGGACACCATGATATGCTGCTGAGTG 1000
334 IuProGlnGluLeuLeuValAsnTyrValSerAspValArgValSerPro 350
1001 AACCACAGAACTACTGTGTAATACTATGTGTGAGCTCCGAGTGAATCC 1050
351 AlaAlaProGluArgGlnGlySerSerLeuThrLeuThrCysGluAla 367
1051 GCAGCCCTGAGAGACAGGAAGGACAGCAGCTCACCCCTGACCTGTAGGC 1100
367 agIuSerSerGlnAspLeuGluIuPheGlnTrpLeuArgGluGluThrGlyG 384
1101 AGAGAGTACGACAGACCTCGAGTTCCAGTGGCTGAGAGAAAGACAGGCC 1150
384 IuValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
1151 AGGTGCTGAAGAGGGGCGCTGCTCAGTTGCATGACCTGAAACCGGAG 1200
401 AlAGlyLysGlyTyrArgCysValAlaSerValProSerIleProGlyLe 417
1201 GCAGAGGCGGCGTATCGCTCGTGGCTGTGTGCCAGCATACCCGAGCT 1250
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTrpMetA 434
1251 GAACCGACACACAGCTGTGTCAACGTGGCCATTTTGGCCCCCTTGGATGG 1300
434 IapheLysGluArgLysValTTrpValLysGluAsnMetValIleuAsnLeu 450
1301 CATTCAGAGAGAGAGAGGTGTGGTGAAGAAATATGATGTTGAATCTG 1350
451 SerCysGluIuLaseGlyHisProArgProThrIleSerTrpAsnValAs 467
1351 TCTTGTGAAGCTTACGGGACACCCCGGCCACCATCTCTCGGAACGTCAA 1400
467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThrL 484
1401 CGGACAGGCAAGTGAACAAAGACCAAGATCCACAGCGAGTCTGAGCACCC 1450
484 euAsnValLeuValThrProGluLeuLeuGluIuPheGlyValGluCysThr 500
1451 TGAATGTCTCTGTACCCCGGAGCTGTTGAGACAGGTGTGAATGGACG 1500
501 AlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluLeuVal 517
1501 GCCTCCAAACGACCTGGGCAAAAACACACAGCATCTCTTCTCGGAGCTGT 1550
517 IAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSerT 534
1551 CAATTTAAACCCCTTCACACAGACTCCAAACACACAGCTGGGCTTCAGCA 1600
534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
1601 CTTCACAGTCCAGTCTCATACAGAGCAACACAGACCTCCACAGAAAGA 1650
551 LysLeuProGluIuProGluSerArgGlyValValIleValAlaValIleVal 567
1651 AAGCTGCCGAGACCGGAGACCGGGGCGTGTGATCGTGGCTGTGATTTGT 1700
567 IcysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheLeuTyrL 584
1701 GTGCATCTGTGTCTGGGCGTGGGCTGGGCGTGTCTCTATTTCTCTATA 1750
```



```

|||||
708 ACATGAAGAGTCACAGGAAGTCACGCTCCCTGTTTCTACCCGACAGAA 757
251 LysValThrPLeuGluValGluProValGlyMetLeuLysGluLysAsp 267
758 AAAGTGTGCTGGAAAGTGAAGCCCGGAGAAATGCTCAAGGAAGGGGACCG 807
267 gValGluIleArgCysLeuAlaAspGlyAsnProProOHISphSerI 284
808 CCGGAAATCAGGTGTTTGGCTGATGGCAACCTCCACACACTTCAGCA 857
284 LeSerLysGlnAsnProSerThrArgGluAlaGluGluIleThrAsn 300
858 TCAGCAAGACAGAACCCACAGCAGGAGGAGAGAGAGAACAAACAAAC 907
301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGly 317
908 GACAAAGGGGTCTGCTGGTGGAGCTCGCCGGAAGGAAACACAGTGGCG 957
317 gTyrgLysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSerG 334
958 CTATGATGTGCAAGCCTGGAGACTGGACACATGATATGCTGCTGATG 1007
334 LnProlGlnLeuLeuValAsnTyrValSerAspValArgValSerPro 350
1008 AACCAAGAGAACTACTGTGAATATGTGTGACGTCGAGTAGATCC 1057
351 AlaAlaProlGlnArgGlnGlnGlySerSerLeuThrLeuThrCysGln 367
1058 GCAGCCCCCTGAGACAGAGAGGACAGCACCCTACCCCTGATGAGGC 1107
367 agLysSerGlnAspLeuGlnPheGlnTrpLeuArgGluIleThrGlyG 384
1108 AGAGAGTAGGACAGACCTGAGTTCAGTGGCTGAGAGAAAGAGACAGCC 1157
384 LnPValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
1158 AGTGTGTGGAAGGGGCGCTGTGCTCAGTGCATGCACTGAAGACGGAG 1207
401 AlaGlyGlyTyrArgCysValAlaSerValProSerIleProlGlyE 417
1208 GCAGGAGGGGCTATGCTGCTGCGTGGCTGTGCCACGATACCGGCT 1257
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProlTrpMet 434
1258 GAACCGACACACAGCTGTCAAGCTGCCATTTTGGCCCCCTTGATGG 1307
434 LpHelysGluArgLysValThrValLysGluAsnMetValLeuAsnLeu 450
1308 CATTCAAGGAGAGAGTGTGGTGAAGAAATGTGTGAATCTG 1357
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTrpAsnValAs 467
1358 TCTTGTGAAGCTCAGGGCCACCCCGCCACCATCTCTCGAGACCTCAA 1407
467 ngIyThrAlaSerGluGlnAspGlnAspProlArgValLeuSerThrL 484
1408 CGGACAGGCAAGTGAACAACCAAGATCCACAGAGTCTGAGCAGCCC 1457
484 euAsnValLeuValThrProlGluLeuGluIleThrGlyValGlyCysThr 500
1458 TGAATGCTCTGAGCCCGGAGCTGTGGAGACAGGTGTGAATCAGCAG 1507
501 AlSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluLeuVal 517
1508 GCCTCCACACACCTGGGCAAAAACACACAGATCTCTCTCGAGCTGGT 1557
517 LAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyLeuSerT 534
1558 CAATTTAAACACCTCAGCACACAGACTCAACACACACAGCTGCGCTCAGCA 1607
534 hSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
|||||

```

```

1608 CTTCACATGCCATGCTCTATACAGAGCCAAACAGACCTCCACAGAGAGA 1657
551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIleVal 567
1658 AAGCTCCCGAGCCCGAGAGCCGGGCGTGGTCATCTGCTGATGATTT 1707
567 LcysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheLeuTyrL 584
1708 GTGCATCTCTGCTCGTGGGGGTGTGGGGCTGTCTCTCATTTCTCTATA 1757
584 yLysGlyLysLeuProCysArgArgSerGlyLysGlnGluIleThrLeu 600
1758 AGAAGGGCAAGCTCCGCTGAGGCGCTCAGGAGAGAGAGATCACGCTG 1807
601 ProProSerArgLysSerGluLeuValValGluValLysSerAspLysLe 617
1808 CCCCCGTCCTGATAGACCGAACTGTAGTGAAGTTAAGTCAGATAAGCT 1857
617 uProlGluIleMetGlyLeuLeuGlnGlySerSerGlyAspLysArgAlaP 634
1858 CCCAGAAAGATGGGCTCTCTGACAGGCGACGCGGTGACAAAGAGGCTC 1907
634 roGlyAspGlnGlyGluLysTyrIleAspLeuArgHis 646
1908 CGGAGACCAAGGAGAGAAATATCATCATGTGAGGCAAT 1945
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH02921
seq_documentation_block:
ID AAH02921 standard; DNA: 3293 BP.
XX
AC AAH02921:
XX
DT 15-JUN-2001 (first entry)
XX
DE Human shear stress-response coding sequence SEQ ID NO: 95.
XX
KW Human; shear stress-response protein; vascular disease;
XX arteriosclerosis; ds.
XX
OS Homo sapiens.
XX
PN W0200125427-A1.
XX
PD 12-APR-2001.
XX
PE 02-OCT-2000; 2000MO-JP06840.
XX
PR 01-OCT-1999; 99JP-0280976.
XX
PA (RYOW ) KYOMA HAKKO KOGYO KK.
XX
PA (NOJI/) NOJIMA H.
XX
PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K,
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
XX
DR WPI: 2001-266308/27.
XX
DR P-PSDB: AAB90798.
XX
PT DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
PS Claim 20; Page 491-496; 678pp; Japanese.
XX
SS The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 3293 BP; 786 A; 916 C; 924 G; 667 T; 0 other:

```

alignment_scores:

Quality: 3327.00 Length: 646
Ratio: 5.166 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916

alignment_block:

US-09-653-961-2 x AAH02921

Align seg 1/1 to: AAH02921 from: 1 to: 3293

```
1 MetGlyLeuProAlaGlyLeuValCysAlaPheLeuLeuAlaAlaCysCysCy 17
|||||
8 ATGGGGCTTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57
|||||
17 sCysProAlaGlyAlaAlaGlyValProGlyGluAlaGluInProAlaProG 34
|||||
58 CTGTCTCGCGCTGCGGGGTGTGCGCGGAGAGCTGAGCAGCGCTGCGGCTG 107
|||||
34 IuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
108 AGCTGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 157
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLys 67
|||||
158 TCCCGAGTCCCAAGGCAACCTCAGCATGTCGACTGCTTTCTGTCACAA 207
|||||
67 sGluLysArgThrLeuLeuLeuPheArgValArgGlnGlyGlnGlnSerG 84
|||||
208 GGGAGAGCGGAGCGCTCATCTTCCGTGTGCGCCAGGCGCCAGGCGCGAGCG 257
|||||
84 IuProGlnGluTrpGlnGluAlaArgLeuSerLeuGlnAspArgGlyValThr 100
|||||
258 AACCTGGGAGGTACGACGACGCGCTCAGCTTCCAGSACAGAGGCGCTACT 307
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgGlyLeuPheLeuCysG 117
|||||
308 CTGGCCCTGACTCAAGTACACCCCAAGAGAGACGCAATCTTCTGTGTGCA 357
|||||
117 nGlyLysArgProAlaGlySerGlnGluTrpArgGlyLeuLeuValArgVal 134
|||||
358 GGGCAAGGCGCCCTGCTGCTCCAGGAGTACGCAATCCAGCTCCGCGCTACA 407
|||||
134 yAlaAlaProGlnGluProAsnLysGlnValAsnProLeuGlyLysProVal 150
|||||
408 AAGCTCCGGAGGAGCAACATCCAGGTCAACCCCTGCGCATCCCTGTG 457
|||||
151 AsnSerLysGluProGlnGluValAlaThrCysValGlyArgAsnGlyTy 167
|||||
458 AACAGTAAAGACCTGAGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 507
|||||
167 rProLysProGlnValLysTrpTrpLysAsnGlyArgProLeuLysGluG 184
|||||
508 CCCCATTCCTCAAGTCACTGTGTACAAAGATGGCGGCTCTGAGAGAGG 557
|||||
184 IuLysAsnArgValHisLysLeuSerSerGlnThrValGlnSerSerGly 200
|||||
558 AGAAGAACCGGGTCCACATTCAGTCCGACAGTGTGAGTGCAGATGCT 607
|||||
201 LeuTrpThrLeuGlnSerLysLeuLysAlaGlnLeuValLysGlnAspGly 217
|||||
608 TTGTACACCTTCGACAGATTTCTGAAAGCAGCAGCTGTGTTAAAGAACAA 657
|||||
217 sAspAlaGlnPheTrpCysGluLeuAsnTrpArgLeuProSerGlyAsnH 234
|||||
658 AGATGCGCCAGTGTACTGTGAGCTCAACATACGGGCTGCCAGTGGAGAAC 707
|||||
234 sMetLysGlnSerArgGluValThrValProValPheTrpProThrGlu 250
|||||
708 ACATGAGAGAGTCCAGGAGATCACGCTCCGTGTTTTCACCGACAGAA 757
|||||
251 LysValTrpLeuGluValGluProValGlyMetLeuLysGlnGlyAspArg 267
|||||
```

```
758 AAAGTGTGCTGGAAGTGAAGCCCGTGGGAATGCTGAAGAGAGGAGACCG 807
267 gValGluLeuLeuArgCysLeuAlaAspGlyAsnProProHisPheSerI 284
|||||
808 CGTGGAAATCAGGTGTGTGCTGATGTGGACACCTCCACACACTCAGCA 857
|||||
284 LeuSerLysGlnAsnProSerThrArgGluAlaGluGluGluThrThrAsn 300
|||||
858 TCAGCAAGCAGAACCCCGACCGAGGAGGAGCAGAGAGAGAGACAAACAC 907
|||||
301 AspAsnGlyValLeuValLeuGluIuProAlaArgGlyLysHisSerGlyAr 317
|||||
908 GACAAAGGGGTCTCGGTGCTGAGGCTGCGCGGAAAGACAGATGGGCG 957
|||||
317 gTrpGlnCysGlnLysLeuAspLeuAspThrMetLysSerLeuLysSerG 334
|||||
958 CTATGAATGTACGGCTGAGAACTTGGACCATATATCCGTGCTGAGTG 1007
|||||
334 IuProGlnGluLeuLeuValAsnTrpValSerAspValArgValSerPro 350
|||||
1008 AACCAAGGAACTACTGTGAATATGTGTGAGCTCCGAGTGAATCC 1057
|||||
351 AlaAlaProGlnLysArgGlnGlySerSerLeuThrLeuThrCysGluAl 367
|||||
1058 GCAGCCCTGAGAGACAGAGAGGAGACAGCTCAGCTGAGCTGTGAGGC 1107
|||||
367 sGluSerSerGlnAspLeuGluIuPheGlnTrpLeuArgGluLysThrGly 384
|||||
1108 AGAGAGTACGACGACCTCGATGTCCAGTGGCTGAGAGAGAGAGAGAC 1157
|||||
384 IuValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
|||||
1158 AGCTCTGGAAGAGGAGGCTGTCTTCACTTCACTGATGATCAACAGGAG 1207
|||||
401 AlaGlyGlyLysTrpArgCysValAlaSerValProSerLysProGlyLe 417
|||||
1208 GCAAGAGGCGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
|||||
417 uAsnArgThrGlnLeuValAsnValAlaLeuPheGlyProProTrpMetA 434
|||||
1258 GAAACCGCACAGCTGCTGCAAGCTGGCCATTTTGGCCCCCTGTGATGG 1307
|||||
434 LaPheLysGluArgLysValTrpValLysGluAsnMetValLeuAsnLeu 450
|||||
1308 CATTCAGAGAGAGAGTGTGTGCTGTAAGAAATATGCTTGTGATATC 1357
|||||
451 SerCysGluAlaSerGlyHisProArgProThrLysSerTrpAsnValAs 467
|||||
1358 TTTGTGAAGCTGAGGAGACCCCGGCGCCACCATCTCTGGAACGTCAA 1407
|||||
467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThrL 484
|||||
1408 CGGCAGCGCAAGTGAACNAGCANAGATCCAGCGAGTCCGTGAGACCC 1457
|||||
484 euAsnValLeuValThrProGluLeuLeuGluThrGlyValGluCysThr 500
|||||
1458 TGAATGCTCTGCTGACCCGAGACCTGTGAGACAGGTGTAATGACAG 1507
|||||
501 AlaSerAsnAspLeuGlyLysAsnThrSerLysLeuPheLeuGluLeuVal 517
|||||
1508 GCTTCACAGACCTGGCAGAAACACACACATCTCTCCGAGGAGGAG 1557
|||||
517 IuAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSer 534
|||||
1558 CAATTTAACACCTCTCACACACATCCAAACCAACACTGGCTCAGCA 1607
|||||
534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
|||||
1608 CTTCACATGCGCAGTCTCATACACAGAGCCAAACAGACCTCCACAGAGA 1657
|||||
551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIleVal 567
|||||
1658 AAGCTGCGGAGCGGAGAGCCGGGCGGTGTCATCGTGGCTGTGATGTG 1707
|||||
```

```
567 ICysIleLeuValIleuAlaValIleuGlyAlaValIleuTyrrhLeuTyrL 584
|||||
1708 GTGCATCTCGTCTGCGGCTGCGGCTGCTCTATTCCTCTATTA 1757
584 yslYsgIlyLysLeuProCysArGArGSerGlyLysGInGlyLeu 600
1758 AGAAGGCAAGCTCCGCTGACAGGCGCTCAGGAAAGCAAGATCAGCTG 1807
601 ProProSerArGlySerGlyLeuValYalGluValLysSerAspLysLe 617
1808 CCCCCTCTCTGTAAGACCAACTGTGTGAAGTTAAAGTCAGATPAAGCT 1857
617 uProGluGluMetGlyLeuLeuGInGlySerSerGlyAspLysArGAlap 634
1858 CCCAGAAAGAGATGGGCTCTCTGACAGGCGACAGCGGTGACAAAGGGCTC 1907
634 roGlyAspGInGlyLysTyrIleAspLeuArGHis 646
1908 CGGAGAGCAAGGAGAGAAATACATCGATCTGAGGCAAT 1945
seq_name: /STD5/9cdata/geneseq/geneseqn-emb1/NA1999.DAT.AA220932
seq_documentation_block:
ID AA220932 standard; cDNA; 1962 BP.
XX AA220932;
XX AC
XX 01-DEC-1999 (first entry)
XX
XX Human MUC18 cDNA sequence, modified to facilitate cloning.
XX
XX prostate cancer; melanoma; cell adhesion; glycoprotein; cloning;
XX metastasis; treatment; detection; diagnostic test; ds.
XX Homo sapiens.
XX
XX OS
XX FH
XX Key Location/Qualifiers
XX primer_bind 1..26
XX FT /*tag= a
XX FT /*note= "Primer BF1"
XX FT primer_bind complement (1931..1955)
XX FT /*tag= b
XX FT /*note= "ER6A primer"
XX
XX PN MO9945392-A1.
XX
XX PD 10-SEP-1999.
XX
XX PE 02-MAR-1999; 99WO-US04850.
XX
XX PR 03-MAR-1998; 98US-0076664.
XX
XX PA (DYEM-) UNIV EMORY.
XX
XX PI Wu G;
XX
XX DR WPI: 1999-540899/45.
XX
XX PT Detection of metastatic prostate cancer, by detection of MUC18
XX expression in prostate cancer cells
XX
XX PS Disclosure: Page 9; 80pp; English.
XX
XX This is the modified nucleotide sequence of MUC18 with primer binding
XX sites (AA220935 and AA220936) to introduce a BamHI site just upstream of
XX the translation start site in order to facilitate cloning.
XX CC The presence of this glycoprotein has been correlated with the ability
XX of melanomas to metastasize. MUC18 is also associated with normal
XX CC vascular tissue, and on the smooth muscle of venules, and it expresses
XX CC sporadically on capillary epithelium.
XX CC The method can be used as a diagnostic test for prostate cancer which
XX has a relatively high potential for metastasis or which has
```

```
CC metastasized. This allows the physician to choose appropriate surgical,
CC chemo- therapeutic or radiation treatment regimens. In addition,
CC antibody specific to MUC18 can be used to prevent metastasis of prostate
CC cancer cells.
XX
XX SO Sequence 1962 BP; 457 A; 559 C; 595 G; 351 T; 0 other;

alignment_scores:
Quality: 3316.00 Length: 645
Ratio: 5.157 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.760

alignment_block:
US-09-653-961-2 x AA220932 ..
Align seg 1/1 to: AA220932 from: 1 to: 1962

1 MetGlyLeuProArgLeuValCysAlaPheLeuAlaAlaCysCysC 17
|||||
10 ATGGGCTTCCAGGCTGTGTGCGCTTCTTGTCTGCGCGCTGCTGCTG 59
17 SCysProArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
60 CTGTCTCGCTGCGGCTGTGCGGAGGCTGAGCAGCTGCGGCTG 109
34 IuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
110 AGCTGTGAGGTGGAAAGTGGGCGAGCACACCTCTGTAAGTGGGCTC 159
51 SerGInSerGInGlyAsnLeuSerHisValAspTrrPheSerValHisL 67
|||||
160 TCCCACTCCCAAGGCAACCTCAGCATGTGACTGTGTTTGTGTCCAAA 209
67 sGlyLysArgThrLeuIlePheArgValArgGInGInGInGInSerG 84
210 GAGAGAGCGAGGCTCATCTTCCGTGTGCGCCAGGCGCCAGGCGCAGG 259
84 IuProGlyGluTyrGluGluArgLeuSerLeuGInAspArgGlyAlaThr 100
|||||
260 AACCTGGGAGACGACGAGCGGCTCAGCTCCAGGACGAGGCGCTACT 309
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgGlyIlePheLeu 117
|||||
310 CTGGCCCTGACTCAAGTCACCCCGCAAGACGCGCATCTTGTGTGCA 359
117 nGlyLysArgProArgSerGInGluTyrArgIleGInLeuArgValTyrL 134
360 GGGCAAGCGCCCTCGGTCCAGAGTACCCCATCCAGCTCCGCGTACA 409
134 ySAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
410 AACCTCCGAGGAGCCAAACATCCAGGTCAACCCCTGGGCAATCCGTGTG 459
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyT 167
460 AACAGTAAGAGACCTGAGAGGTGCTACTGCTGTAGAGAGACAGGTA 509
167 rProIleProGlnValIleTrrTyrLysAsnGlyArgProLeuLysGlu 184
510 CCCCATCTCTCAAGTCATCTGTATCAAGATGCGGCTCTGAAGAGAG 559
184 IuLysAsnArgValHisIleGInSerSerGlnThrValGluSerGly 200
560 AGAAGAACCGGCTCCATTCATGTCGCCAGACTGTGAGTCAAGTGTGT 609
201 LeuTyrThrLeuGInSerIleLeuLysAlaGlnLeuValLysGluAsp 217
610 TTGTACACCTTCACAGATATCTGAAGGCAAGCTGTGTTAAAGAACAA 659
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsn 234
660 AGATGCCAGTTTACTGTAGTCAACTACCGGCTGCCAGTGGGAACC 709
```



```

234 isMetLysGluSerArgGluValThrValProValPheTyrProThrGlu 250
|||||
710 ACATGAAAGAGTCCAGGGAGACACCTGCTTTTCTTACCACAGAA 759
251 LysValThrPheGluValGluProValGlyMetLeuLysGluLysPhe 267
|||||
760 AAAGTGTGGCTGGAAGTGGAGCCCTGGGAATGCTGAAGGAAGGGACCG 809
267 GValGluLeuArgCysLeuAlaAspGlyAsnProProProHisSerI 284
|||||
810 CCGTGAATACAGGTGTTGGCTGATGGCAACCTCCACACACTTCAGCA 859
284 LeSerLysGluAsnProSerThrArgGluAlaGluGluThrThrAsn 300
|||||
860 TCAGACACAGAACCCCGACAGGGGAGGAGGAAGAGACACACAC 909
301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGly 317
|||||
910 GACAAAGGGGTCTGCTGGTGTGGAGCTGCCCGGAAGACACAGTGGCG 959
317 GTTGTGTCGAGGTCGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 960
960 CTATGAATGTCAGGCTGGAACCTTGACACCATGATGCTGCTGATG 1009
334 LysProGluGluLeuLeuValAsnTyrValSerAspValArgValSer 350
|||||
1010 AACCAAGAGAACTACTGCTGAACATGTCGACGTCGACGTCGACGTC 1059
351 AlaAlaProGluArgGluGluGluSerSerLeuThrLeuThrCysGlu 367
|||||
1060 GCACCCCTGAGAGACAGGAAGGACAGCCTACCTGACCTGTGAGGC 1109
367 agLysSerSerGluAsnLeuGluPheGlnThrLeuArgGluThrGly 384
|||||
1110 AGAGAGTAGCCAGACCTCCAGTCCAGTGGAGAGGAAGAGACGAGC 1159
384 InValLeuGluArgGlyProValLeuGluHisAspLeuLysArgGlu 400
|||||
1160 AGGTGCTGGAAGGGGGCCCTGCTTCAAGTGCATGACCTGAAAGGG 1209
401 AlAGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGly 417
|||||
1210 GCAGGAGGGCGCTATGCTGCGGCGTGTGCCAGATACCGGCCCT 1259
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTyrP 434
|||||
1260 GAACCCGACACAGCTGTGCAAGCTGGCCATTTTGGCCCCCTGTGAT 1309
434 IapheLysGluArgLysValTyrValLysGluAsnMetValLeuAsn 450
|||||
1310 CATTCAGAGAGAGAGAGGTGGGTGAAGAATATGCTGTGATCTG 1359
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTyrAsnVal 467
|||||
1360 TCTTGTGAAGCGTCAGGACCCCGGCCACACATCTCTGGAACGTCAA 1409
467 ngLysThrAlaSerGluGluAspGluAspProGluArgValLeuSer 484
|||||
1410 CGGACGCGCACTGAACAAGACCAAGATCCACAGAGATCCTGAGCAC 1459
484 euAsnValLeuValThrProGluLeuLeuGluThrGlyValGluCys 500
|||||
1460 TGAATGTCTCTGTGACCCCGAGCTGTGGAGACAGGTGTAATGCAC 1509
501 AlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluVal 517
|||||
1510 GCCTCAAGAGACCTGGGCAAAAACCGACATCTCTTCCGAGGCTGGT 1559
517 IasLeuThrThrLeuThrProAspSerAsnThrThrGlyLeuSerI 534
|||||
1560 CAATTAAACACCCCTCACACAGACTCCACACCAACACAGTGGCTCAG 1609

```

```

534 InSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGlu 550
|||||
1610 CTTCCACAGTCCAGTCTCTATACAGAGCCACACACACTCCACAGAGA 1659
551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIle 567
|||||
1660 AAGCTGCCGGAGCCCGAGAGCCGGGGCGTGTGATCTGCTGTGATG 1709
1567 LysIleLeuValLeuAlaValLeuGluAlaValLeuTyrPheLeu 584
|||||
1710 GTGCATCTGTGCTCTGGGCTGTGGGCTGTCTCTTATTTCTCTTANA 1759
584 LysLysGlyLysLeuProCysArgArgSerGlyLysGluLeuLeu 600
|||||
1760 AGAAGGCAACCTGCCGTCAGGCGCTCAGGGAAGCAGGAGATCAGCTG 1809
601 ProProSerArgLysSerGluLeuValValGluValLysSerAspLys 617
|||||
1810 CCCCCGTCTGTAGACCGAATGTTGAGTTAAGTCAAGTCAAGTCA 1859
617 uProGluGluMetGlyLeuLeuGluGluGluSerSerGlyAspLysArg 634
|||||
1860 CCCAGAGAGATGGGCTCTCTGCAAGGACACACGCTGACAAAGAGGCT 1909
634 roGlyAspGluGluGluLysTyrIleAspLeuArg 645
|||||
1910 CGGAGACCAAGGAGAGAAATACATCATCTGAAAG 1944
seq_name: /sids5/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA661554
seq_documentation block:
ID AA661554 standard; cDNA, 524 BP.
XX
XX AA661554;
AC
XX
XX 19-FEB-2001 (first entry)
DT
XX
XX Genetic suppressor elements which is derived from Mel-CAM cDNA.
DE
XX
XX Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM;
KW beta-integrin; invasive melanoma; retrovirus vector; solid tumour;
KW angiogenesis; ss.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX MO200058501-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 24-MAR-2000; 2000WO-US07807.
PF
XX
XX 26-MAR-1999; 99US-0126479.
PR
XX
XX (WIST-) WISTAR INST.
PA
XX
XX Herlyn M, Salyamoorthy K;
PI
XX
XX WPI; 2000-594646/56.
DR
XX
XX Trans-recoverable packaging deficient retrovirus vectors for the
PT treatment of solid tumors -
PT
XX
XX Claim 19; Fig 3A; 84pp; English.
PS
XX
XX AAC61554-74 represent genetic suppressor elements which exhibit an
CC anti-melanoma effect. The polynucleotides are derived from a portion
CC of cDNA corresponding to the cell adhesion protein Mel-CAM or
CC beta-integrin. The genetic suppressor elements either inhibit expression
CC or induce hyper-expression of Mel-CAM or beta-integrin. Expression of
CC these proteins are known to be correlated with survival and growth of
CC invasive melanomas. The specification describes methods for inhibiting
CC a phenotype associated with diseased cells. The methods use gene

```

CC suppressor elements for inhibiting such phenotypes. The method involves
CC using a trans-recoverable packaging deficient retrovirus vector to
CC deliver a plurality of polynucleotides to a population of target cells.
CC The methods are used for treatment of solid tumours and for inhibiting
CC angiogenesis in solid tumour cells.

XX Sequence 524 BP; 124 A; 139 C; 166 G; 95 T; 0 other;

alignment-scores: Length: 174
 Quality: 888.00
 Ratio: 5.133
Percent Similarity: 99.425 Percent Identity: 98.276

alignment-block:
US-09-653-961-2 x AAC61554 ..

Align seg 1/1 to: AAC61554 from: 1 to: 524

```
327 ThmetileserleuSerGluProGlnGluLeuValAsnTyra 343
|||||
1 ACCATGATATCGCTGAGTGAACACAGAACTACTGTGAACTATGT 50
343 lSerAspValArgValSerProAlaAlaProGluArgGlnGlySerS 360
|||||
51 GTCGTACGCTCCGAGTCCCGCAGCCCTGAGAGACAGAAAGGAGCA 100
360 erleuThleuThrCyseGluAlaGluSerSerGlnAspLeuGluPheGln 376
|||||
101 GCTTACCTGACCTGCTGAGGACAGAGTAGCCAGACCTGAGTTCCAG 150
377 TtleuArgGluGluThrGlyGluValLeuGluArgGlyProValLeuG 393
|||||
151 TGGCTGAGAGAAAGACAGACCAGTGTGAAAGGGGCGCTGTCTCA 200
393 nleuThAspLeuLysArgGluAlaGlyGlyTyraGcysValAla 410
|||||
201 GTTGCAATGACTGAACCGGAGGAGGAGGCGCTATCGCTGCGCT 250
410 erValProSerIleProGlyLeuAsnArgThrGlnLeuValAsnValAla 426
|||||
251 CTTGCCCCGACATACCGCGCTGAACCGCACACACCTGTGTAACCTG 300
427 llePheGlyProProIlePheAlaPheLysGluArgLysValTyrVal 443
|||||
301 ATTTTGGCCCCCTTGATGTCATTCAAGAGAGAGAGTGTGGCTAA 350
443 sGluAsnMetValLeuAsnLeuSerCysGluAlaSerGlyHisProArgP 460
351 AGAGAAATATGGTGTGAATCTGTGTGTAACCTGAGGCGACCCCGCGC 400
460 roThrIleSerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAsp 476
401 CCACCAATCTCTGGAAGCTCAAGCGCAGCGCAAGTGAACACCAAGAT 450
477 ProGluArgValLeuSerThrLeuAsnValLeuValThrProGluLeu 493
451 CCAACAGGAGTCTGAGCAACCTGGAATGCTCTGTGACCCCGAGCTGTT 500
493 uGluThrGlyValGluCysThr 500
501 GGAGACAGCTGTGAAATGACG 522
```

seq_name: /stids/gcdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC61560

seq_documentation_block:

ID AAC61560 standard; CDNA: 422 BP.

XX AAC61560;

XX 19-FEB-2001 (first entry)

XX Genetic suppressor elements which is derived from Mel-CAM cDNA.

XX Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM;
KW beta-Integrin; invasive melanoma; retrovirus vector; solid tumour;
KW angio genesis; ss.

OS Synthetic.
OS Homo sapiens.

PN WO200058501-A2.

XX 05-OCT-2000.

PF 24-MAR-2000; 2000WO-US07807.

XX 26-MAR-1999; 99US-0126479.

XX (WIST-) WISTAR INST.

XX Herlyn M, Satyamoorthy K;

XX WPI; 2000-594646/56.

XX Trans-recoverable packaging deficient retrovirus vectors for the
PT treatment of solid tumors -

XX Claim 19; Fig 10C; 84pp; English.

XX AAC61554-74 represent genetic suppressor elements which exhibit an
CC anti-melanoma effect. The polynucleotides are derived from a portion
CC of cDNA corresponding to the cell adhesion protein Mel-CAM or
CC beta-integrin. The genetic suppressor elements either inhibit expression
CC or induce hyper-expression of Mel-CAM or beta-integrin. Expression of
CC these proteins are known to be correlated with survival and growth of
CC invasive melanomas. The specification describes methods for inhibiting
CC a phenotype associated with diseased cells. The methods use gene
CC suppressor elements for inhibiting such phenotypes. The method involves
CC using a trans-recoverable packaging deficient retrovirus vector to
CC deliver a plurality of polynucleotides to a population of target cells.
CC The methods are used for treatment of solid tumours and for inhibiting
CC angiogenesis in solid tumour cells.

XX Sequence 422 BP; 103 A; 125 C; 118 G; 76 T; 0 other;

alignment-scores: Length: 140
 Quality: 698.00
 Ratio: 4.986
Percent Similarity: 100.000 Percent Identity: 99.286

alignment-block:
US-09-653-961-2 x AAC61560 ..

Align seg 1/1 to: AAC61560 from: 1 to: 422

```
504 AspleuGlyLysAsnThrSerIleuPheLeuGluLeuValAsnLeuTh 520
1 GACCTGGGCAAAAAACACGACATCTCTTCCGTGGAGCTGGCAATTAC 50
520 rThrIleuThrProAspSerAsnThrThrThrGlyLeuSerThrSerThra 537
51 CACCTTCACACGAGCTCCACACACACCTGCTGAGCACTTCCACTG 100
537 lAserProHisThrArgAlaAsnSerThrSerThrGluArgLysLeuPro 553
101 CCACTCTCTATACCAAGCCCAAGACACCTCCACAGAGAGAAAGCTGCCG 150
554 GluProGluSerArgGlyValValIleValAlaValIleValCysIlele 570
151 GAGCCGAGAGAGCCGGCGGTGTCATCGTGGCTGTGATTTGTGCATCT 200
570 uValLeuAlaValLeuGlyValValLeuTyrPheLeuTyrLysGlyL 587
201 GGTCTGCGCGTGTGCGGCGCTGCTCTATTTCCCTAATTAAGAGAGGCA 250
```

587 yslLeuProCysArgArgSerGlyLysGlnGluIleThrLeuProProSer 603
|||||
251 AGTCGCGCTGCGAGCGCTCAGGAGACAGAGATCAGCTGCCCGCTCT 300
|||||
604 ArgLysSerGluLeuValValGluValLysSerAspLysLeuProGlu 620
|||||
301 CGTAAAGACCGAACTTGTAGTTGAAGTTAACTCAGATTACCTCCAGAGA 350
|||||
620 uMetGlyLeuGlnGlnLysSerGlyAspLysArgAlaProGlyAspG 637
|||||
351 GATGGGCTCTCTCGAGGCGACGGGTGACAAAGGGCTCGGAGAGACC 400
|||||
637 lngLysGluLysTyrIleAsp 643
|||||
401 AGGAGAGAAATACATCAT 420
|||||

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA00314

seq_documentation_block:

ID AAC00314 standard; cDNA; 441 BP.

AC AAC00314;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 312.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR P-PSDB; AAG00308.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 312; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

XX Sequence 441 BP; 72 A; 149 C; 135 G; 81 T; 4 other;

alignment_scores:

Quality: 682.00 Length: 136
Ratio: 5.128 Gaps: 1
Percent Similarity: 97.794 Percent Identity: 97.794

alignment_block:

US-09-653-961-2 x AAC00314 ..

Align seg 1/1 to: AAC00314 from: 1 to: 441

1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaIaCysCysCY 17
|||||
34 ARGGGGCTTCCAGGCTGTGCGCTCTTCTGCTCGCGGCTGTGCTG 83
|||||
17 scYsProArgValAlaIaGlyValProGlyGluAlaGlnProAlaProG 34
|||||
84 CTGTCTCGCGCGCGGGGTGTCCCGGAGAGGCTGACGAGCTCGGCTG 133
|||||
34 luleuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
134 ACCTGTGTGAGGTGGAAGTGGGACACAGCCCTTCTGAAGTGGGCTC 183
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLys 67
|||||
184 TCCAGTCCCAAGGCAACCTCAGCCATGTGCGATGTTTCTGTCCAA 233
|||||
67 sglLysArgThrLeuIlePheArgValArgGlnGlyGlnLysGlnSerG 84
|||||
234 GAGAGAGCGAGCGCTCATCTTCCGTGTCGCGACGAGGCGCAGAGCG 283
|||||
84 lUpProGlyGluTyrGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
284 AACCTGGGAGTACGAGCAGCGGCTCAGCCCTCCAGAGACAGGGGCTACT 333
|||||
101 leuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCys.G 117
|||||
334 CTGGCCCTGACTCAAGTCACCCCAAGACGAGCGCATCTTCTGTGCAN 383
|||||
117 lngLysArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyr 133
|||||
384 GTNRMAAGCGCCCTCGTGTCCAGAGATACCGCATTCAGCTCGCTTAC 433
|||||
134 LysAla 135
|||||
434 AAAGCT 439
|||||

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA061562

seq_documentation_block:

ID AAC61562 standard; cDNA; 400 BP.

AC AAC61562;

DT 19-FEB-2001 (first entry)

DE Genetic suppressor elements which is derived from Mel-CAM cDNA.
XX
KW Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM;
KW beta-integrin; invasive melanoma; retrovirus vector; solid tumour;
KW anglogenesis; ss.

OS Synthetic.

OS Homo sapiens.

PN M0200058501-A2.

PD 05-OCT-2000.

PF 24-MAR-2000; 2000MO-US07807.

PR 26-MAR-1999; 99US-0126479.

PA (WIST-) WISTAR INST.

XX

```
PI Herlyn M, Salyamoorthy K;
XX WPI; 2000-594646/56.
DR
XX
XX Trans-recoverable packaging deficient retrovirus vectors for the
PT treatment of solid tumors -
XX
XX Claim 19; Fig 11B; 84pp; English.
PS
XX AC61554-74 represent genetic suppressor elements which exhibit an
CC anti-melanoma effect. The polynucleotides are derived from a portion
CC of cDNA corresponding to the cell adhesion protein Mel-CAM or
CC beta-Integrin. The genetic suppressor elements either inhibit expression
CC or induce hyper-expression of Mel-CAM or beta-Integrin. Expression of
CC these proteins are known to be correlated with survival and growth of
CC invasive melanomas. The specification describes methods for inhibiting
CC a phenotype associated with diseased cells. The methods use gene
CC suppressor elements for inhibiting such phenotypes. The method involves
CC using a trans-recoverable packaging deficient retrovirus vector to
CC deliver a plurality of polynucleotides to a population of target cells.
CC The methods are used for treatment of solid tumours and for inhibiting
CC angiogenesis in solid tumour cells.
XX
XX Sequence 400 BP; 95 A; 98 C; 133 G; 74 T; 0 other;

Alignment_scores:
Quality: 655.00 Length: 132
Ratio: 5.038 Gaps: 0
Percent Similarity: 98.485 Percent Identity: 95.455

Alignment_block:
US-09-653-961-2 x AAC61562 ..

Align seg 1/1 to: AAC61562 from: 1 to: 400

315 serGLYARITyrgLUcYgInGLyLeuAsPLeuAsPThMeILSerLe 331
|||||
3 AGTGGCGCTATGATGATGTCAGCGCTTGACACTTGACACCATATATCGCT 52
|||||
331 uLeuSerGIuPProGLInGLuLeuValAsnTyRValSerAsPValArgV 348
|||||
53 GCTGAGTGAACACAGACGACTACTGCTGAACTATGCTGTGACGTCGAG 102
|||||
348 aLSeRProAlaAProGLuArgGLInGLySerSerLeuThr 364
|||||
103 TGAgtCCCGAGCGCTGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 152
|||||
365 CyAGlUALAGlUSeRSeRGLInAsPLeuGLuPheGLInTRPLeuArgGLu 381
|||||
153 TGTGAGAGCAGAGAGTAGCCAGACCTCGAGTTCCAGTGGCTGAGAGAGA 202
|||||
381 uThrGLyGLInValLeuGLuArgGLyProValLeuGLInLeuHISAsPLeuL 398
|||||
203 GACAGAGACAGAGTGTGAAAGGGGGCTGTCTTCAGTTGCATGACAGACTGA 252
|||||
398 ySArGLuAlaGLyGLyGLyTYrARgCySAValAlaSeRValProSeRILE 414
|||||
253 AACGGAGAGAGAGAGCGGCTATCGCTGCTGGCGTCTGTGCCACACATA 302
|||||
303 CCGGGCGCTGAACCGCACACAGCTGTGTCAAGCTGGCCATTTTGGCCCCC 352
|||||
415 ProGLyLeuAsnARgThrGLInLeuValAsnValAlaIlePheGLyProP 431
|||||
431 oTPMeAlaPheLySGLuArgLySValTRPValLySGlUAsnMet 446
|||||
353 TTGATGTGCAATTCAGAGAGAGAGAGAGGTGTGTGTAAGAGACAAATATG 398
|||||

seq_name: /std5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABA83083
seq_documentation_block:
ID ABA83083 standard; DNA; 2402 BP.
XX
```

```
AC ABA83083;
XX
XX 08-FEB-2002 (first entry)
DT
XX
XX Lutheran blood group protein (BCAM) ovarian tumour marker gene, #5.
DE
XX
XX epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
KW
XX
XX identification; serous cystadenoma; borderline serous tumour;
KW
XX
XX serous cystadenocarcinoma; mucinous cystadenocarcinoma;
KW
XX
XX mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
KW
XX
XX undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
KW
XX
XX adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
KW
XX
XX immune response pathway; cell proliferation regulation; protein folding;
KW
XX
XX membrane localised; secreted; therapeutic target; cytostatic;
KW
XX
XX gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
XX WO200175177-A2.
XX
XX 11-OCT-2001.
XX
XX 03-APR-2001: 2001WO-US10947.
XX
XX 03-APR-2000: 2000US-194336P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
XX
XX P-P-SDB; ABB50259.
XX
XX WPI: 2001-626450/72.
XX
XX Detecting and identifying ovarian tumor, identifying increased risk for
PT developing ovarian cancer, and determining effectiveness of ovarian
PT cancer treatment, by measuring expression level of ovarian tumor marker
PT gene -
XX
XX Claim 22; Page 54; 140pp; English.
XX
XX The invention relates to methods for diagnosing and prognosing ovarian
XX tumours in an individual via the detection and measurement of the
XX expression of an ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,
XX ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
XX ABA83181 and ABA83183). The methods of the invention are useful for
XX detecting an ovarian tumour in a patient, for identifying an individual
XX at increased risk for developing ovarian cancer, in prognostic tests for
XX assessing the relative severity of ovarian cancer, in tests for
XX monitoring a patient in remission from ovarian cancer and in tests for
XX monitoring disease status in a patient being treated for ovarian cancer.
XX The methods can additionally be used to identify a particular tumour as
XX being an ovarian tumour (i.e., an epithelial ovarian tumour selected from
XX serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
XX mucinous cystadenoma, borderline mucinous tumour, mucinous
XX cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
XX clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
XX tumour. The ovarian tumour marker genes of the invention were identified
XX using SAGE (serial analysis of gene expression) and were found to be
XX overexpressed in a broad variety of ovarian epithelial tumour cells
XX relative to normal ovarian epithelial cells. The marker genes are
XX implicated in immune response pathways, in the regulation of cell
XX proliferation and in protein folding, and many of these are membrane-
XX localised or secreted. In addition to their use as diagnostic and
XX prognostic markers, the ovarian tumour marker genes or their encoded
XX proteins may be used as therapeutic targets for the treatment and
XX prevention of ovarian cancer. Sequences ABA83081-ABA83122, ABA83180,
XX ABA83182 and ABA83184 represent the ovarian tumour marker genes of
XX the invention.
XX
XX Sequence 2402 BP; 426 A; 844 C; 742 G; 390 T; 0 other;
XX
```


326TACCACTGACTCCAGGGGCC.....CTGTGCT 357
103 uThrGlnValThrProGlnAspGluArgIlePheLeuCys.....GlnG 118
118 LysArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrLys 134
408 CAGGGGGCGGACACTGCTGAGGCACACTGGGGCTCAACGCTGTTCGA 457
135 AlaProGlnGluProAlaIleGlnValAspProLeuGlyIleProValAs 151
458 AAGCAGAGGGCCACTGAGGTCTCCCAACAAAGGACACTGCTGTGAT 507
151 uSerLysGluProGlnGluValAlaThrCysValGlyArgAsnGlyTyrP 168
508 GGAGGACTCTGCCAGAGATCGCACACTTCAACAGCCGGAAGGAGACC 557
168 rolleProGlnValIleTyrTyrLysAsnGlyArgProLeuLys..... 182
558 CGGCCCCCAAGATCACGTGTATCGCAACAGCGGCGCTGAGAGTGGCC 607
183 GluGluLysAsnArgValHisIleGlnSerSerGlnThrVal...GluSe 198
608 GTAGAGATGAACCCAGAGGGCTACATGACACGCCGACGCTCGGAGAGC 657
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
658 CTGGGGCTGCTCTCCCTCACAGACCCTTACCTGCGGCTCCGCAAG 707
215 LuAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProser 231
708 ATGACGAGAGCCGACCTTCCACTGCGCGCCGACACAGCTCTGCCAG 757
232 GlyAsnHisMetLys...GluSerArgGluValThrValProValPheTyr 247
758 GGCGGCGGCGCGCTGAGACAGCCCACTTCCACCTCACCTCCACTA 807
247 rProThrGluLysVal.....TyrLeu.....GluValGluProValG 260
808 TCCACAGGACGACGCTGCTGTGGTGGGACCCGCTCCACCCGAGAG 857
260 LysMetLeuLysGluLysAspArgValGluIleArgCysLeuAlaAspGly 276
858 GCTGGGTACGAGGGGTGACTGTCAGCTGCTGCGGGGGAGCGC 907
277 AsnProProHisPheSerIleSerLysGlnAsnProSerThrArgGly 293
908 AGCCCAAGCCGAGATATACGCTTTCGCG.....CTTCAGGA 945
293 uAlaGlnGluGluThrThrAsnAspAsn.....GlyValLeuValLeuG 308
946 TGAGCAGAGGAAATGCTGAATGATCTGAGGGGAACTTGACCTGG 995
308 LuProAlaArgLysGluHisSerGlyArgTyrGlyCysGlnGlyLeuAsp 324
996 AGGAGATACCCCGGGGCGGACGCTATGCTGACAGAGTGGAGAT 1045
325 LeuAspThrMetIleSerLeu...LeuSerLupProGlnGluLeuLeuVal 340
1046 TACGAGGGGACGATGACGTCTCCAAAGAGCTGAGAGCTGCCGT 1095
340 LAsnTyrValSerAspValArgValSerProAlaAlaProGluArgGln 357
1096 GGCCTATCTGAGCCCTGAGAGTCAGC.....G 1124
357 LuGly.....SerSerLeuThrLeuThrCys 365
1125 AGGGAGAGGCTCTTCTTACCTTAACAGCAGTGCAGCTGGAACCTG 1174
366 GluAlaGluSerSerGlnAspLeuGluPheGlnThrLeuArgGlnGlu 382
1175 TCCGTGACGGCTGCGCCACCCCTGACGTGAGCAAGACTCCAC 1224

382 rGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysA 399
1225 T.....CCCTGGGCGATGGCCCAATGCTGTCTGATGTTATACCT 1268
399 rGluAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
1269 TCGATTCGAATGACCTACGTATGTGAGGCTCCCTGCCACAGTCCCG 1318
416 GlyLeuAsnArgThrGlnLeuValAlaIlePheGlyProProTr 432
1319 GTCTTCAGCCGACCCAGAACTTCACTGCTGCGCGGCTCATCAGCCCA 1368
432 pMetAlaPheLysGluArgLys.....ValTyrValLysG 444
1369 GCTAAAGACAGCGGAATATAGCCCAAGCAGATGACACTGG...AGGG 1415
444 LuAsnMetValLeuAsnLeuSerCysGluAlaSerGlyHisProArgPro 460
1416 AAGGAGACGAAGTCACACTCATCTGCTGCGCGGCTCATCAGACCC 1465
461 ThrIleSerTyrAsnValAsnGlyThrAlaSerGlnGlnAspGlnAspPr 477
1466 AAATCGAGTGGAGACCAATTGGGG...GGCAGCCCGGACAGCAATCC 1512
477 oGlnArg.....ValLeuSerThrLeuAsnValLeuValThrProG 491
1513 CGAGCGGACAGGTTGGGTGGAGCAGCTCTGACCTGAAAGTACACAGC 1562
491 LuLeuLeuGluThrGlyValGlyCysThrAlaSerAsnAspLeuLys 507
1563 CCTGAGCCGCGATGCGATCTCTGTGAACTCCACCCGAGGAGAAC 1612
508 AsnThrSerIleLeuPheLeuGluLeuValAsnLeuThrThrLeuThr 524
1613 AAGCGCAGTCTTCCACTTGGCGCGGTAGC..... 1645
524 oAspSerAsnThrThrGlyLeuSerThrSerThrAlaSerProHisT 541
1645 1645
541 hArgAlaAsnSerThrSerThrGluArgLysLeuProGluProGluSer 557
1646CCCAAGACTCCAG 1660
558 ArgGlyValValIleValAlaValIleValCysIleLeuValLeuAlaVal 574
1661 GCTGAGTGGCCGTATGCGCTGCGCTGAGCTGAGCTCTGCTCT 1710
574 LLeuGlyAlaValLeuTyrPheLeuTyrLysGlyLysLeuProCysA 591
1711 CGTCTGCTGCTCTCTACTGCGTGAAGCAGCAAGGGGGC...CCCTGCT 1757
591 rArgSerGlyLysGlnGluIleThrLeuProProSerArgLysSerGlu 607
1758 GCCGACGACGGCGGAGAGAGAGGGGCTCCGCCCA.....GGGAG 1798
608 LeuValValGluValLysSerAspLysLeuProGlnGluMetGlyLeu 624
1799 CCAGGGCTGAGACCACTCGGGGTGAGCAACAGAGACGACCGCTCT 1848
624 uGlnGly.....SerSerGlyAspLysArgAlaProGlyAspGlnGlyG 639
1849 CATGGAGAGTGCCTCGGAGAGACGAGGCTGGCACGGGGCTTGGAG 1898
639 LuLys 640
1899 ACAG 1903

seq.name: /STD5/gc9data/geneseq/geneseqn-emb1/NA2001B.DAT:AA572009
seq_documentation_block:
ID AA572009 standard: cDNA: 2402 BP.

alignment_scores:

Quality: 633.00 Length: 123
Ratio: 5.189 Gaps: 0
Percent Similarity: 99.187 Percent Identity: 97.561

alignment_block:
US-09-653-961-2 x AAC61555

Align seg 1/1 to: AAC61555 from: 1 to: 373

```
378 LeuArgGluGluThrGlnValLeuGlnAlaArgGlyProValLeuGlnLe 394
|||||
3 CTAGAGAAAGAGACAGACCAGCTCTGAAAGGGGCGCTGTCTCAGATT 52
|||||
394 uHisAPLeuLysArgGluAlaGlyGlyTyrArgCysValAlaSerV 411
|||||
53 GCATGACCTGAAACGGAGCGACGCGCTATCCGCTGCGGCTCG 102
|||||
411 aProSerIleProGlyLeuAsnArgThrGlnLeuValAsnValAlaIle 427
|||||
103 TGCCCAAGCATACCAGCGCTGAACCGCACACAGCTGGTCAAGCGCCATT 152
|||||
428 PheGlyProProTrrPMeTAlaPheLysGluArgLysValTrrVallySG 444
|||||
153 TTTGGGCCCCCTTGATGTCATCAAGAGAGAGAGGTGGGTGAAGA 202
|||||
444 uAsnMetValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProT 461
|||||
203 GAATATAGGTGTTGAATCTGTCTGTGAAGCGTGAAGGCGACCCCGCCA 252
|||||
461 hTrrLeuSerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspPro 477
|||||
253 CCATTCCTCTGAAACGTCAGCGCAAGTGAACAGCAAGATCA 302
|||||
478 GlnArgValLeuSerThrLeuAsnValLeuValThrProGluLeuGln 494
|||||
303 CACGAGGTCTCTGACGACCCCTCAATGTCTCTGTACCCCGAGCTGTGGA 352
|||||
494 uThrGlyValGluCysThr 500
|||||
353 GACAGGTGTGAATGCACG 371
|||||
seq_name: /SID55/gcdata/geneseq/geneseqn-emb1/NA001A.DAT:ABA09243
seq_documentation_block:
ID ABA09243 standard; cDNA; 2617 BP.
XX
AC ABA09243;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human Lutheran blood group glycoprotein homologue cDNA, SEQ:1019.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoietic regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemokinesis; chemokinesis; thrombolytic; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiaesthetic; antiarrhythmic; haemostatic; antiarteriosclerotic;
KW cytosolic; osteopathic; vasotropic; cardiac; virucide; antibacterial;
KW antifungal; vulnery; antidiabetic; ss.
XX
XX Homo sapiens.
XX OS
XX WO200157188-A2.
XX PN
XX 09-AUG-2001.
XX PD
```

```
PF 05-FEB-2001; 2001WO-US03800.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
XX PPI: 2001-457740/49.
XX P-PSDB: ABB11999.
XX
XX Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PS e.g. arthritis and cancer.
XX
XX Claim 1: Page 859-860; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX invention also relates to vectors and recombinant host cells comprising a
XX nucleotide of the invention, methods of producing the novel polypeptides,
XX or polypeptides in a sample, and methods of identifying compounds which
XX bind to polypeptides of the invention. Although novel, many of the
XX polypeptides of the invention have homology to known proteins, and hence
XX giving an insight into their probable biological activities, and hence
XX potential therapeutic applications. The polypeptides of the invention may
XX have various activities, including cytokine, cell proliferation or cell
XX differentiation activities; stem cell growth factor activity;
XX haematopoietic regulatory activity; tissue growth activity;
XX immunomodulatory activity; activin- or inhibin-related activities;
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX thrombolytic activities; receptor or ligand activities; or may be
XX involved in oncogenesis, cancer cell proliferation and metastasis.
XX depending on their biological activities, polypeptides and nucleotides of
XX the invention are useful for preventing, treating or ameliorating medical
XX conditions, e.g., by protein or gene therapy. Such conditions include
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX proliferative retinopathy, atherosclerosis, coronary heart disease,
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX vascular growth. Polypeptides involved with tissue regeneration and
XX healing (e.g., of burns, incisions and ulcers), while those with
XX immunomodulatory activities may be used in the treatment of viral,
XX bacterial and fungal infections in addition to immune disorders.
XX polypeptides with growth factor activity may be used in cell cultures to
XX promote cell growth. For example, such polypeptides may be used to
XX manipulate stem cells in culture to give rise to neuroepithelial cells
XX that can be used to augment or replace cells damaged by illness.
XX autoimmune disease or accidental damage. The polypeptides and nucleotides
XX may also be used in the diagnosis of the above conditions, and in drug
XX screening techniques. The present sequence represents a cDNA encoding a
XX novel human polypeptide of the invention.
XX
XX Sequence 2617 BP; 533 A; 1018 C; 626 G; 439 T; 1 other:
SQ
alignment_scores:
Quality: 606.00 Length: 640
Ratio: 1.717 Gaps: 26
Percent Similarity: 55.156 Percent Identity: 29.844
alignment_block:
US-09-653-961-2 x ABA09243
XX
XX Align seg 1/1 to: ABA09243 from: 1 to: 2617
XX
XX 2 GlyLeuProArg...LeuValCysAlaPheLeuLeuAlaCysCysCy 17
XX ||| ||||| ||::: ||| ||||| |||||
XX 56 GGGGCCCGCGCTGTCTGTCTGCGAGTCTGCTGCGCGG..... 97
```

```
17 sCysProArGValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
98 .....CACCCAGATGCCAAGCGAGGTGGCTGTGTCTGTATCCCC 137
34 LuLeuValGluValGluValSerThrAlaLeuLeuLysCysGlyLeu 50
138 CGTGTGTGAGGTGATGCGAGGAAAGTGTTCATTTCTGGACTGC..... 181
51 SerGlnSerGlnGlyAsnLeuSerHis.....ValAspTrpPhe..... 63
182 ...ACCCCTACGGGAGACCCAGACCATTTATGTCTGGAATGTCTTAC 228
64 .....SerValHisLysGluLysA 70
229 CGACCGCTCGGAGCTCGCCCGCTAGCTCGGTGAGTGCAGAGGCT 278
70 rGThrLeuIlePheArgValArgGlnGlyGlnGlyGlnSerGluProGly 86
279 CTGAGCTTCAGAGTCAATGACAGACACCGGGGCGCAGTCTCCCCA... 325
87 GluTyrGluGlnArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLe 103
326 .....TACCACTGAGCTCCACAGGGCGC.....CTGTGCT 357
103 uThrGlnValThrProGlnAspGlnArgIlePheLeuCys.....GlnG 118
358 GGTGAGGCGCAGGTGGCGACGACGAGCTGTGTGCTGTGAGG 407
118 LysAspArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrLys 134
408 CAGGGGCGGACGAGCTGCTGAGGGCGCTCGGGCTCAACGTGTTGCA 457
135 AlaProGluGlnProAsnIleGlnValAsnProLeuGlyTleProValAs 151
458 AACCCAGAGCGCACTGAGTCTCCCCACAAAGAGACACTGTCTGTGAT 507
151 nSerLysGluProGluGlnValAlaThrCysValGlyArgAsnGlyTyrP 168
508 GAGAGACTCTGCCAGAGATGCCACACGACCAACAGCCGGAAGCAACC 557
168 rGLeuProGlnValIleTyrTyrLysAsnGlyArgProLeuLys..... 182
558 CGGCCCCCAAGATCAGGTGTATGCAAGCGGAGCGCTGGAGTGGCCC 607
183 GluGluLysAsnArgValHisIleGlnSerSerGlnThrVal...GluSe 198
608 GTAGAGATGAACCAAGAGGCTACATGACACCGCACGCTCCGGGAGGC 657
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
658 CTGGGCGCTGTCTCTCCACACGACCTCTACCTCGCGCTCCGCACAGG 707
215 LuAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
708 ATGACGAGAGCGCACAGCTTCCACTGCGCGCCACTACAGCTCCGCGG 757
232 GlyAsnHisMetLys...GluSerArgGluValAlaThrAlaProValPheTyr 247
758 GGCAGCGACGGCGCTGGACAGCCCACTTCCACCTCACCTGCACACA 807
247 rProThrGluLysVal.....TrpLeu.....GluValGluProValG 260
808 TCCACAGAGCAGCTGACATCTGGGTGGGCGCCCGCTCCACCCACAG 857
260 LysMetLeuLysGlnGluLysAspArgValGluIleArgCysLeuAlaSpLy 276
858 GTGGGTACGGGAGGTGACACTGTCCAGCTGTCCGGGGGAGGAGCG 907
277 AsnProProHisPheSerIleSerLysGlnAsnProSerThrArgG 293
908 AACCCAGCGCGGAGTATACGCTTTTCCG.....CTTCAAGA 945
293 uAlaGluGlnGluTyrThrAsnAspAsn.....GlyValLeuValLeuG 308
946 :.....: 946 :.....:
308 LuProAlaArgLysGluHisSerGlyArgTyrGluCysGlnGlyLeuAsp 324
996 AGGAGTAGACCCGGGCGCAGAGGGGACCTATGTGCTGCAAGTGAAGAT 1045
325 LeuAspThrMetIleSerLeu...LeuSerGluProGlnGluLeuValA 340
1046 TACAGCGCGCAGATGACGTGACGCTTCCAGAGCGCTGAGCTGGGCT 1095
340 LasnTyrValSerAspValArgValSerProAlaAlaProGluArgGln 357
1096 GGCCTATCTGGACCCCTGGAGCTCAGC.....G 1124
357 LuGlySerSerLeuThrLeu.....ThrCys 365
1125 AGGGAAGGTGCTTCTTCTTACCTTAACACAGAGGCGAGCTGTGAATGC 1174
366 GluAlaGluSerSerGlnAspLeuGluPheGlnTyrPheArgGluGluTh 382
1175 TCCGTGACAGGCGTGCACCCCTGCTACCTGAGCAACAGAGACTCCAC 1224
382 rGlyGlnValLeuGluArgGlyProValProValLeuHisAspLeuLysA 399
1225 T.....CCCTGGGCGATGCGCCCTGCTGTGCTGCTATACACT 1268
399 rGluAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
1269 TCGATTCCAAATGGCACATATGATGAGGCTCTCCCTCCACAGTCCG 1318
416 GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTr 432
1319 GTCCTGAGCGGCCCAAGAACTTCACTGCTGTGTCCAAAGCTGCCAGA 1368
432 PheAlaPheLysGluArgLys.....ValTrpValLysG 444
1369 GCTAAAGACAGCGAATAGAGCCCAAGGAGATGCCAGCTGG...AGG 1415
444 LuAsnMetValLeuAsnLeuSerCysGluAlaSerGlyHisProArgPro 460
1416 AAGGAGCAGAGTACACTCATCTGCTGCCCGCCGACATCCAGACAGC 1465
461 ThrIleSerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspPr 477
1466 AAACCTACACTGGAGCAATTGGG...GGCAGCCCGCAGAGCCAAATCC 1512
477 GlnArg.....ValLeuSerThrLeuAsnValLeuValThrProG 491
1513 CGAGCGCAGGCTGGGTGAGACACTCTCTGACCTGAAGTGAACAGCG 1562
491 LuLeuLeuGluThrGlyValGluCysThrAlaSerAsnAspLeuGlyLys 507
1563 CCTGAGCGCGGAGTGGCATCTCTGTGAAGCTTCAAC..... 1600
508 AsnThrSerIleLeuPheLeuGluLeuValAsnLeuThrThrLeuThrP 524
1601 .....CC 1602
524 AspSerAsnThrThrThrGlyLeuSerThrSerThrAlaSerProHisT 541
1603 CCAGGGAACAAGCGCATGCTTCCACTTCCGACACCGTGAACCCACAG 1652
541 hrArgAlaAsnSerThrSerThrGluArgLysLeuProGluProGluSer 557
1653 CC.....TCCCAG 1660
558 ArgGlyValValIleValAlaValIleValCysIleLeuValLeuAlaVa 574
1661 GCTGGAATGGCGCTCATGCGCGTGCAGCGTGGGCTCTGCTGCT 1710
574 LLeuGlyAlaValLeuTyrPheLeuTyrLysGlyLysLeuProCysA 591
```

```
1711 CGTCGTCGTCTCTCTACTGCTGAGACGCAAGGGGCG...CCCTGCT 1757
591 rgatgsergilyglinglu 597
|||||:|||||:
1758 GCCGCCAGCGCGGGAGAG 1777

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA61559

seq_documentation_block:
ID   AAC61559 standard; cDNA: 297 BP.
XX
AC   AAC61559;
XX
DT   19-FEB-2001 (first entry)
XX
DE   Genetic suppressor elements which is derived from Mel-CAM cDNA.
XX
KW   Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM;
KW   beta-integrin; invasive melanoma; retrovirus vector; solid tumour;
KW   angiogenesis; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200058501-A2.
XX
PD   05-OCT-2000.
XX
PE   24-MAR-2000; 2000WO-US07807.
XX
PR   26-MAR-1999; 99US-0126479.
XX
PA   (WIST-) WISTAR INST.
XX
PI   Herlyn M, Salyamoorthy K;
XX
DR   WPI: 2000-594646/56.
XX
PT   Trans-recoverable packaging deficient retrovirus vectors for the
PT   treatment of solid tumors -
XX
PS   Claim 19; Fig 10B; 84pp; English.
XX
CC   AAC61554-74 represent genetic suppressor elements which exhibit an
CC   anti-melanoma effect. The polynucleotides are derived from a portion
CC   of cDNA corresponding to the cell adhesion protein Mel-CAM or
CC   beta-integrin. The genetic suppressor elements either inhibit expression
CC   or induce hyper-expression of Mel-CAM or beta-integrin. Expression of
CC   these proteins are known to be correlated with survival and growth of
CC   invasive melanomas. The specification describes methods for inhibiting
CC   a phenotype associated with diseased cells. The methods use gene
CC   suppressor elements for inhibiting such phenotypes. The method involves
CC   using a trans-recoverable packaging deficient retrovirus vector to
CC   deliver a plurality of polynucleotides to a population of target cells.
CC   The methods are used for treatment of solid tumours and for inhibiting
CC   angiogenesis in solid tumour cells.
XX
SQ   Sequence 297 BP; 69 A; 69 C; 103 G; 56 T; 0 other;

alignment_scores:
      Quality: 501.00      Length: 98
      Ratio: 5.165      Gaps: 0
Percent Similarity: 98.980      Percent Identity: 96.939

alignment_block:
US-09-653-961-2 x AAC61559 ..
Align seg 1/1 to: AAC61559 from: 1 to: 297

356 Glingluyserserleuthrleuthrcysglualagluserserlins 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 CAGGAGGACGACGACCTCACCCTGACCTGAGAGCAGAGAGTACGACAGA 52

372 pleugluhpheglntprleuarglgluirlhrglgluValleugluarag 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 CCHGAGTCCAGTGGCTGAGAGAGACAGACACGAGCTGGAAAGG 102
389 lProValleuglnleuInsAspLeuLysArgGluIaGlyGlyTyr 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 GGCTGTGCTTCAGTTGCATGACCTGAACAGGAGGAGGAGCGCGTAT 152
406 ArgCysValAlaSerValProSerIleProGlyLeuAsnArgThgIle 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 CGTCGCTGGGCTGTGCGCCAGCATACCGGCTGACCGACACAGCT 202
422 uValAsnValAlaIlePheGlyProProIleMetAlaPheLysGluArgL 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 GGTCAAGCTGGCCATTTTGGCCCCCTTGATGGCATTCAGAGAGAGA 252
439 ySValTTPValLysGluAsnMetValLeuAsnLeuSerCysGlu 453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 AGTGTGCGGTGAAGAGATAATGTTGAATCTGCTTGTGA 296

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF68399

seq_documentation_block:
ID   AAF68399 standard; cDNA: 4235 BP.
XX
AC   AAF68399;
XX
DT   12-APR-2001 (first entry)
XX
DE   Human lung tumour protein related nucleotide sequence SEQ ID NO:317.
XX
KW   Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW   lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW   cytostatic; antisense inhibition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200100828-A2.
XX
PD   04-JAN-2001.
XX
PE   30-JUN-2000; 2000WO-US18061.
XX
PR   30-JUN-1999; 99US-0346492.
XX
PR   15-OCT-1999; 99US-0419386.
XX
PR   17-DEC-1999; 99US-0468687.
XX
PR   30-DEC-1999; 99US-0476300.
XX
PR   06-MAR-2000; 2000US-0519642.
XX
PR   22-MAR-2000; 2000US-0533077.
XX
PR   10-APR-2000; 2000US-0546259.
XX
PR   27-APR-2000; 2000US-0560406.
XX
PR   05-JUN-2000; 2000US-0589184.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T, Banpur CS, Iodes MJ, Fanger GR, Vedlack TS, Carter D;
PI   Retter MW, Mannion J;
XX
DR   WPI: 2001-071488/08.
XX
PT   Lung tumor-associated proteins and the nucleic acids that encode them,
PT   useful for preventing, diagnosing and treating lung cancer -
XX
PS   Example 1; Page 244-245; 436pp; English.
XX
CC   The present invention describes immunogenic portions of lung tumour-
CC   associated proteins (I) and the nucleic acids (NAs) that encode them.
CC   (I) have cytostatic activity and can be used in gene therapy, antisense
CC   inhibition and in vaccines. The NAs and the lung tumour-associated
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with their inappropriate expression,
CC   especially lung cancers. For example, the NAs may be administered to
```


[illegible]


```

380  GlnGluThrGlyValValLeuGluArgGlyProValLeuGlnLeuHisAs 396
      ::::: ||||| ::::: ::::: :::::
1168  GAAAGACATCAGG....CTGCATCTAGCCG....TCATTTCTCAG 1205
      ::::: ::::: ::::: :::::
396  PleuLysArgGluValGlyGlyTyArgCysValAlaSerValProS 413
      ::::: ::::: ::::: :::::
1206  TCTTCATTACAGATCGTGAACATATGTCTCCGAACCTGCTGTCAG 1255
      ::::: ::::: ::::: :::::
413  SerLeuProGlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGly 429
      ::::: ::::: ::::: :::::
1256  AGCTTAAGAGACTTAAGAAAAGAGACTCATTCATCTCATTTGAGAGGC 1305
      ::::: ::::: ::::: :::::
430  ProProTrpMetAlaPheLysGluArgLysValTrpValLysGluAsn 446
      ::::: ::::: ::::: :::::
1306  AATCCT....CAATTAATAATGACAAAGAAATCATGATCCAGTCGACT 1349
      ::::: ::::: ::::: :::::
446  ValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleS 463
      ::::: ::::: ::::: :::::
1350  ATCTAAACATTAATCTGCCATGTGGAGGTTTCCAAAGCCAGCCATTC 1399
      ::::: ::::: ::::: :::::
463  SerTrpAsnValAsnGlyThrAlaSer.....GluGlnAspGlnAsp 476
      ::::: ::::: ::::: :::::
1400  ARTGGACAATTACTGCGCAGTGGAGCGCTATAACCAACAGAGCATCT 1449
      ::::: ::::: ::::: :::::
477  ProGln.....ArgValLeuSerThrLeuAsnValLeuValThrPr 490
      ::::: ::::: ::::: :::::
1450  CCTTAATATTAAATGCGAGTATTACT....AAATATTATCATTTCCCC 1493
      ::::: ::::: ::::: :::::
490  OGluLeuLeuGluThrGlyValGluCysThrAlaSerAsnAspLeuGly 507
      ::::: ::::: ::::: :::::
1494  TGAAGAGAAATGTACA...TTAACTTGACACAGCAAAACCAACTGAG 1540
      ::::: ::::: ::::: :::::
507  YAsnThrSerIleLeuPheLeuGluLeuValAsnLeuThrThrLeuThr 523
      ::::: ::::: ::::: :::::
1541  GAACAGTAATACCTCTGAATGTCTCGCTATAAGTATT..... 1578
      ::::: ::::: ::::: :::::
524  ProAspSerAsnThrThrThrGlyLeuSerThrSerThrAlaSerProH 540
      ::::: ::::: ::::: :::::
1579  CCAGAA.....CCAGAA.....CA 1586
      ::::: ::::: ::::: :::::
540  sThrArgAlaAsnSerThrSerThrGluArgLysLeuProGluProGluS 557
      ::::: ::::: ::::: :::::
1587  CGATGAGGCGACGAGATAGATGATAAACAGAA...GAAAGGTGAAATG 1633
      ::::: ::::: ::::: :::::
557  eArgGlyValValIleValAlaValIleValCysIleLeuValLeuAla 573
      ::::: ::::: ::::: :::::
1634  ACCAGCAAAATCAATTGTGGAAATGTGTGCTCCCTGCTGCTGCC 1683
      ::::: ::::: ::::: :::::
574  ValLeuGlyAlaValLeuTyThrPheLeuTyThrLys..... 585
      ::::: ::::: ::::: :::::
1684  CTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1733
      ::::: ::::: ::::: :::::
586  .....GlyLysLeu.ProCysArgArgSer 593
      ::::: ::::: ::::: :::::
1734  ATCAAAACATGTAAACAGACCTCGGTATATGAGAAACAAACAAAGT 1783
      ::::: ::::: ::::: :::::
594  GlyLysGlnGluIleThr...LeuProProSerArgLysSerGluLeuVal 609
      ::::: ::::: ::::: :::::
1784  TAGAAGAAACATCAACAAACAGAGCGTAAGAGAAACAGTCTCTAGT 1833
      ::::: ::::: ::::: :::::
609  ValGluValLysSer.....AspLysLeuProGluGluMetGlyLeu 623
      ::::: ::::: ::::: :::::
1834  TGTCCAGAGATTAATAATCATATATAGAACCAATTTGAACATGTAAC 1883
      ::::: ::::: ::::: :::::
seq_name: /sgn2_6/ptodata/2/1na/PCtus_COMB.seq: PCT-US93-00031-22
seq_documentation_block:
; Sequence 22, Application PC/TUS9300031
; GENERAL INFORMATION:
; APPLICANT: Osborn, Laurelee
; APPLICANT: Benjamin, Christopher D.
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
; IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1

```

```

? NUMBER OF SEQUENCES: 23
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Allegretti & Witcoff, Ltd.
? STREET: 10 South Wacker Drive
? CITY: Chicago
? STATE: IL
? COUNTRY: US
? ZIP: 60606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/00031
? FILING DATE: 19930112
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: McNicholas, Janet M.
? REGISTRATION NUMBER: 32,918
? REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (312) 715-1000
? TELEFAX: (312) 715-1234
? INFORMATION FOR SEQ ID NO: 22:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1941 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: double
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1941
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 73..1941
? FEATURE:
? NAME/KEY: sig_peptide
? LOCATION: 1..72
? PCT-US93-00031-22

alignment_scores:
Quality: 244.50 Length: 683
Ratio: 0.743 Gaps: 32
Percent Similarity: 48.170 Percent Identity: 21.523

alignment_block:
US-09-653-961-2 x PCT-US93-00031-22 ..
Align seg 1/1 to: PCT-US93-00031-22 from: 1 to: 1941

31 ProAlaProGluLeuValGluValAlGlySerThrAlaLeuLeu 47
   ||||| ::::: ::::: ::::: :::::
91 CTTGACCCCGG..ATTGCTGCTCAGATGTGAGACTCAGTCATGTTGAC 137
   ||||| ::::: ::::: ::::: :::::
47 sCysGlyLeuSerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheS 64
   ||||| ::::: ::::: ::::: :::::
138 ATGTAGTGTCAATGCGCTGTAATCC.....CAATCTTCTCTGCG.... 177
   ||||| ::::: ::::: ::::: :::::
64 eValHisLysGlnLysArgThrLeuIlePheArgValArgGlnGln 80
   ||||| ::::: ::::: ::::: :::::
178 .....AGAACCCAGATGACAGACCCCTGTGAGCGGAG 210
   ||||| ::::: ::::: ::::: :::::
81 GlyInSerGluProGlyGlyTyThrGlnGlnArgLeuSerLeuGlnAspAr 97
   ::||| ::::: ::::: ::::: :::::
211 GTGAGAGTGAG.....GGGAC 227
   ::||| ::::: ::::: ::::: :::::
97 gGlyAlaThrLeuAlaLeuThrGlnValThrProGlnAspGluArgIleP 114
   ::||| ::::: ::::: ::::: :::::
228 CAATTCACGCTGACCTGAGCCCTGTGATGATTGAGAACGAAACATCTT 277
   ::||| ::::: ::::: ::::: :::::
114 heLeuCysGln.....GlyLysArgProArgSerGlnGluTyArg 127

```

[illegible]

```

1091 CGGATATGAATCTCTAGAGAACAAAGATTGGAAATGACCTTCATCCCT 1144
399 .....ArgLnaIaGlyGlyTYrArGcysValaIaSerVal..... 411
      :::::::::::::: :::: :::: :::: ::::
1141 ACCATTGAAGATACGTGAAAGAGCTCTGTGTTGTCAGGCTAACTACATAT 1190
412 .....ProSerIleProGlyLeuAsnArgThrcInL 422
      :::::::::::::: :::: :::: :::: ::::
1191 TGATGACATGAATTCGAAACCCAAACAAAGGACAGAGTACCCAAACACTTT 1240
422 euValaSnValaIaIlePheGlyProProThrPheMetaIaPheLysGluArg 438
      ||||| ||||| |||
1241 AAGTCATATTTGCC.....CCC.....ACAGATACA 1266
439 LysValaTrpValLysGluAsnMetValLeu.....As 449
      ||| |||| :::: :::: ::::
1267 ACCGCTTGTCGACCCCTTCCTCATCTCAGGAAAGGACAGTTCTGTGAA 1316
449 nleuSerCysGluIaIaSerCylisThraArgProThrIleSerTrpAsnV 466
      :::::::::::::: :::: :::: ||| ||| ||| ||| ::::
1317 TATGACATCTTGAGCCAGGAGGCTTCTGCTGCCGAAAATCCTGTGGAGC. 1365
466 aIaSnGlyThrIaSerGluIaAspIaSnppProGluArgValLeuSer 482
      ||| |||| :::: ::::
1366 .....AGGCAGCTCCCTAACGGGAGCTACAG 1392
483 ThrLeuAsnValLeuValThrProGluLeuGluThrGlyValGlu.. 498
      |||| :::: :::: :::: :::: ::::
1393 CCTCTTGTGAGAAATGCAACTCTCACCTTAATTTCTCAAAAATGAGAA 1442
499 .....CysThrIaSerAsnAspLeuGlyLysAsnThrS 510
      ||| :::: |||| :::: ::::
1443 TTCTGGGGTTATTATATGTGAAGAAATTAACAGGCTGGAAAGACGAA 1492
510 erIleuPheLeuGluLeu.....ValaSnLeuThr 520
      :::: :::: ||| :::: ::::
1493 AGGAACTGCAATTAATTAATTCCAAGTTACTCCAAAAGACATTAACCTTACA 1542
521 ThrLeuThrProAspSer.....AsnThrThrThrGlyLeuSe 533
      :::: :::: :::: ::::
1543 GCTTTCCTTCTGTGAGATGTCCAAAGAAAGAGACACTGTCAATCATCTCTG 1592
533 rThrSerThrIaSerProHisThr.....ArgIaIa 544
      ||| |||| :::: :::: ::::
1593 TACATGTGAAATGTGCCAAGAACATGATTAATCTGAAAGAAAAGCGG 1642
544 snSerThrSerThrGluArgLys..... 551
      :::: :::: |||
1643 AGCAGAGAGACACAGTACTAAATCTTAAGATGGCGCTATACCATCCGA 1692
552 .....LeuProGluIuProGluSerArgIyVa 560
      :::: :::: ||| |||| ::::
1693 AAGGCCAGTTGAAGGATGGCGGAGTATATGAATGTGAATCTAAAAAACA 1742
560 IVal..... 561
      |||
1743 AGTTGGCTCAAAATTAGAAAGTTTAACACTTGATGTTCAAGAAAGAAA 1792
562 .....IleValaIaValaIleIleValLys 568
      :::: :::: |||
1793 ACACAAAGACTATTTTTCCTCGAGCTTCTCGCTCTATTTTGCAATCC 1842
569 IleLeuValLeuIaValLeuGlyIaIaValaLeuThrPheLeuTYrLys.. 584
      |||| :::: :::: ||| ||| |||
1843 TCCTTAATATATACCTGCAATTTGGAAATGATATATTACTTTGCAAGAAAAGC 1892
585 .....LysGlyLysLeuProCysArgArgSerGlyLysGluInL 598
      |||| :::: :::: ||| ::::
1893 CAACATGAAGAGGTCATATGATCTTGTAGAAAGCACAGAAATCAAAATG 1941
seq_name: /sgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-184-009-144
seq_documentation_block:

```

```
Sequence 144, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
APPLICANT: Proletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESS: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
ZIP: 10036
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/184,009
FILING DATE: 19-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 144:
SEQUENCE CHARACTERISTICS:
LENGTH: 2434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-184-009-144

alignment_scores:
Quality: 242.00 Length: 616
Ratio: 0.910 Gaps: 23
Percent Similarity: 43.182 Percent Identity: 21.104

alignment_block:
US-09-653-961-2 x US-08-184-009-144 ..
Align seg 1/1 to: US-08-184-009-144 from: 1 to: 2434

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValIph 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
748 AGCGTGCACATGTCATGCGACAGACGACCTCATCTATTTCATATGCTAC 797
107 rProGlnAspGluArgGlyLeuPheLeuGlnGlyLeuArgProArgSer. 123
|:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
798 AAGAAATGACACAGACGACCTCAATGTGAACCCAGAACCCAGAGAGT 847
124 ....GlnGlyArgGlyLeuGlnLeuArgValTyrLeuAlaProGluGlu 138
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
848 CAGGCGCGAGTATCATGATCATCTGAATGCTCTATGCGCCGAGATGCC 897
139 ProAsnIleGlnValAsnProLeuGlyIleProValAsnSerLeuGluPr 155
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
898 CCCACCATTT.....TCCCCTTAACACACATCTTACAGATCAGGGGAAA 941
155 oGlnGluValAlaThrCysValGlyArgAsnGlyTyrProIleProGln 172
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
942 TGTGAACCTCTCTGCGACGACGACCTTAC.....CCACTGCACAGT 985
172 aIleTyrTyrLeuAsnGlyArgProLeuGlnGluLeuAsnArgVal 188
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
986 ACTCTGTGTTGTCTATGCG.....ACT 1008
189 HisIleGlnSerGln.....ThrValGluSe 198
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1009 TTCCAGATCCACCCAGACGCTCTTATCCCAACATCACTGATGAATAA 1058
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuValAlaGlnLeuValysG 215
|||||:|||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|
1059 TAGTGATCTTATACGTGCGCAAGCC..... 1083
215 LuAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1084 .....CATACTCA 1092
232 GlyAsnHisMetLysGluSerArgGluValThrValProValPheTyrPr 248
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1093 GACACTGGCCCTCATAGACCCAGCTCAGCAGATCAGTCTATGCGAGA 1142
248 oThrGluValThrLeuGlnValGluProValGlyMetLeuLysGluG 265
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1143 GCCACCCAAACCCCTTATCATCACAGCAACACTCCACCCCGTGAGGATG 1192
265 LysPargValGluIleArgCysLeuAlaAspGlyAsnProProHis 281
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1193 AGGATGCTGTAGCCTTACCTGTGACCTGATTCAGAACACACACTAC 1242
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluGlu 298
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1243 CTGTGTGGTGAATATATCAAGAGCCTCCCGTCACTGCGGCGCGCT 1292
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1293 GTCCATATACAAACAGAGACCTCACTATCACTACAGTCAAGCAATATG 1342
315 erGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1343 TAGGACCTTATGAGTGT...GGAATCCAGAACAAATTAAGTGTGAC... 1386
332 LeuSerGluProGlnGluLeuLeuValAsnTyr...ValSerAspValAr 347
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1387 CACAGCGACCCAGTATCTGATATGCTCTATGCGCCAGACAGCCACC 1436
347 gValSerProAlaIleProGluArgGlnGluGlySerSerLeuThrLeu 364
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1437 CATTTCCCTCATACACATATTCCTGCGAGGCTGATACCTGACCTCT 1486
364 hrCysGluAlaGluSerSerGlnAspLeuGluPheGlnTyrLeuArgGlu 380
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1487 CCTGCCATGACGCTCTAACCCACCTGACACAGTATCTGTGGTATGAT 1536
381 GluThrGlnValLeuGluArgGlyProValLeuGlnHisAspLe 397
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1537 .....GGAGACATCCAGACACACACAAAGAGCTCTTATCTCCACAT 1580
397 uLysArgGluAlaGlyGlyTyrArgCysValAla..... 409
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1581 CACTGAGAGAAGACGCGAGCTCTATCTGCGAGCCCAATATCTACGCCA 1630
409 ..... 409
1631 GTGGCAGACAGAGCTACAGTCAAGCAATACAGTCTGTGGAGAGCTG 1680
409 ..... 409
1681 CCCAAGCCCTTCATCTCAGACACAACTCCAAACCCGTGAGAGACAAGA 1730
409 ..... 409
1731 TGTGTGGCTTACCTGTGACCTGAGGCTCAGAAACAACTACTCTGT 1780
410 .....SerValPro.SerIleProGlyLeu..... 417
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```



```

1084 .....CATAACTCA 1092
232 GYAAsnHISMetLysGluSerArgGluValIThrValProValPheTyrPr 248
      :::::::::::::::::::::
1093 GACACTGGGCTCATTATGAGCACACAGTACAGTACAGTCTATGACGA 1142
      :::::::::::::::::::::
248 oThrGluValThrLeuGluValAlaGluProValGlyMetLeuLysGlu 265
      :::::::::::::::::::::
1143 GCCACCCAAACCTTCATCATCCAGCAACAACCCCGGAGAGATG 1192
      :::::::::::::::::::::
265 LysArgValGluIleArgCysLeuAlaAspGlyAsnProProProHis 281
      :::::::::::::::::::::
1193 AGATGCTGTAGCCTTAACCTGTGAACTGAGATTGACAAACAACCTAC 1242
      :::::::::::::::::::::
282 PheSerIleSerLysGluAsnProSerThrArgGluAlaGluGluIth 298
      :::::::::::::::::::::
1243 CTGTGGTGGTAAATATATACAGAGCTCCGGTCAGTCCAGGCTGAGCT 1292
      :::::::::::::::::::::
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
      :::::::::::::::::::::
1293 CTCGAATGACACAGGACCTCACTCACTCACTCACTCACTCACTCACT 1342
      :::::::::::::::::::::
315 eArgLysArgTyrGluCysGluGlyLeuAspLysPheThrMetIleSerLeu 331
      :::::::::::::::::::::
1343 TAGGACCTTATGAGTGT...GGAATCCAGACAAATTAAGTGTGAC... 1386
      :::::::::::::::::::::
332 LeuSerGluProGluGluLeuLeuValAsnTyr...ValSerAspValAr 347
      :::::::::::::::::::::
1387 CACAGGACCCAGTATCTGAAATGCTCTATGCTGATGCTGATGCTGAT 1436
      :::::::::::::::::::::
347 gValSerProAlaAlaProGluValArgGluGlySerSerLeuThrLeu 364
      :::::::::::::::::::::
1437 CATTTCCCTTCATATACCTATTACCTCCAGGGGTAACTCACTGCTCT 1486
      :::::::::::::::::::::
364 hrcYsGluAlaGluSerSerGlnAspLeuGluPheGlnThrPheAspGlu 380
      :::::::::::::::::::::
1487 CCTGCGATGACAGGCTCTAACCACCTGACAGTATCTTGGCTGATGAT 1536
      :::::::::::::::::::::
381 GluThrGlyGluValLeuGluArgGlyProValLeuGluLeuHisAspLe 397
      :::::::::::::::::::::
1537 .....GGGAACATCCAGCAACACACACAGAGCTCTTTATCTCAACAT 1580
      :::::::::::::::::::::
397 uLysArgGluAlaGluGlyTyrArgCysValAla..... 409
      :::::::::::::::::::::
1581 CACTGAGAAAGACACAGGAGCTCTATACCTGCAGGCCAAATACACAGCA 1630
      :::::::::::::::::::::
409 ..... 409
1631 GTGGCCACAGCAGACTACAGTACAGACATACAGTCTGTGGAGAGCTG 1680
      :::::::::::::::::::::
409 ..... 409
1681 CCCAAGCCTCATCTCCAGCAACAACTCAAACCCGTGGAGAGCAAGA 1730
      :::::::::::::::::::::
409 ..... 409
1731 TGCTGTGGCTTACCTGTGAACCTGAGGCTCAGAACACAACCTACCTGT 1780
      :::::::::::::::::::::
410 .....SerValPro.SerIleProGlyLeu..... 417
      :::::::::::::::::::::
1781 GGTGGTAAATGGTACAGAGCTCCAGTACGTCCAGGCTGAGCTGTGC 1830
      :::::::::::::::::::::
417 ..... 417
1831 AATGGCAACAGACCTCACTATTCATGTACAAAGAAATAGCAGCAAG 1880
      :::::::::::::::::::::
418 .....AsnArgThrGlnLeuValAsnValAlaIlePhe 428
      :::::::::::::::::::::
1881 AGCCTATGATGTGGAATCCAGAACTCAGTGTGCAAAACCGAGTGCAC 1930
      :::::::::::::::::::::
429 GlyProProThrPheAla..... 434
      :::::::::::::::::::::
1931 CAGTCACCTGATGCTCTATGAGGCGGACACCCCATCATTTCCCC 1980

```

```

435 .....PheLysGluArgLysValThrValLysGluAsn 446
      :::::::::::::::::::::
1981 CCAGACTGCTCTTACCTTTCGGAGCCGAC..... 2010
      :::::::::::::::::::::
446 eValLeuAsnLeuSerCysGluAlaAspArgLysProAlaProThrIle 462
      :::::::::::::::::::::
2011 .....CTCAACCTCTCTCCACTCGGCTCTTAACCCATATCCCGCAGTAT 2055
      :::::::::::::::::::::
463 SerThrAsnValAlaGluThrAlaSerGluGlnAspGlnAspProGlnAr 479
      :::::::::::::::::::::
2056 TCTTGGCGATCAATGAGTA.....CCGACGA 2084
      :::::::::::::::::::::
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGlu 495
      :::::::::::::::::::::
2085 ACACACACAAGTCTCTTATCGCCAAATACAGCCAAATATATACGCGA 2134
      :::::::::::::::::::::
495 hrcGluValGluCysThrAlaSerAsn..AspLeuGlyLysAspThrSer 510
      :::::::::::::::::::::
2135 CC...TATGCTGTGTTTGTCTTACTGCTGCTGCTGCTGCTGCTGCT 2181
      :::::::::::::::::::::
511 IleLeuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerAs 527
      :::::::::::::::::::::
2182 ATAAGTC..... 2187
      :::::::::::::::::::::
527 nThrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAla 544
      :::::::::::::::::::::
2188 .AAGAGCATCAGAGTCTCTCATCTGAGAACTCTCTGCTCTCAGCTG 2236
      :::::::::::::::::::::
544 snsrThrSerThrGluArgLysLeuProGluProGluSerArgLysVal 560
      :::::::::::::::::::::
2237 GGGGCACT..... 2244
      :::::::::::::::::::::
561 ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuLys 577
      :::::::::::::::::::::
2245 .....GTGGCATCATGATGATGAGTCTGCTGT.....GGGT 2276
      :::::::::::::::::::::
577 aValLeuThrPheLeuTyrLysGlyLysGlyLeuProCysArgArg 592
      :::::::::::::::::::::
2277 TGTCTGATATAGTTTATCTCAGCA...ATTCTGAGCCCGG 2319
      :::::::::::::::::::::
seq_name: /cgn2.6/prodata/2/lna/sb_comb.seq:us-08-460-736-144
seq_documentation_block:
? Sequence 144, Application US/08460736
? Patent No. 5265189
? GENERAL INFORMATION:
? APPLICANT: Paoletti, Enzo
? APPLICANT: Tartaglia, James
? APPLICANT: Cox, William I.
? TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
? NUMBER OF SEQUENCES: 217
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Curtis, Morris & Safford
? STREET: 530 Fifth Avenue
? CITY: New York
? STATE: NY
? COUNTRY: USA
? ZIP: 10036
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/460,736
? FILING DATE: 02-JUN-1995
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/184,009
? FILING DATE: 19-JAN-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Frommer, William S.

```



```

347 gValserProAla1a1ProGluArgGluGlySerSerLeuThrLeuT 364
      ::::::::::::::::::::
1440 CATTTCCCTCCCTACACCTATTACCGTCCAGGCGTAACCTCACCCCTCT 1489
364 hrcysGluAlaGluSerSerGlnAspLeuGluPheGlnTyrLeuArgLu 380
      ::::::::::::::::::::
1490 CCTGCCATCAGCCCTTACCCACCTGCACAGTATTCTTGGTGATGTGAT 1539
381 GluThrGlyGlnValLeuGluValArgGlyProValLeuGlnLeuHisAspLe 397
      ::::::::::::::::::::
1540 .....GGGACATCCACACACACACACAGAGCTCTTATCTCCACAT 1583
397 uLysArgGluAlaGlyGlyGlyTyrArgCysValAla..... 409
      ::::::::::::::::::::
1584 CACTGAGAAAGAACAGCGACTCTATACTGCCAGGCCAATAACTAGTACGCCA 1633
409 ..... 409
1634 GTGGCCACAGCAGACTACAGTCAAGACAAATCAGACTCTCGCGGAGCTG 1683
409 ..... 409
1684 CCCAAGCCCTCCATCTCCAGCAACAACCTCCAAACCGTGAGAGACAAAGGA 1733
409 ..... 409
1734 TGCTGTGGCCTTCACCTGTGAACCTGAGGCTCAGAACACACACCTTACTGT 1783
410 .....SerValPro.Ser1LeProGlyLeu..... 417
      ::::::::::::::::::::
1784 GGTGGGTAATGTGTGAGAGCTCCAGTCACTGCCAGGCTGCAGAGCTGTCC 1833
417 ..... 417
1834 AATGGCAACAGAGCCCTCACTATTCAATGTTCACAAGAATGACCGCAAG 1883
418 .....AsnArgThrGlnLeuValAsnValAlaIlePhe 428
      ::::::::::::::::::::
1884 AGCCTATGTGTGGAATCCAGAACTCAGTGAAGTGCACAAACCGCAGTAC 1933
429 GlyProThrPheMetAla..... 434
      ::::::::::::::::::::
1934 CAGTACACCTCGATGTCCTATGAGCGCGGACACACCCCATCATTTCCCCC 1983
435 .....PheLysGluArgLysValIleTyrValLysGluAsnM 446
      ::::::::::::::::::::
1984 CCAGACTCGTCTTACCTTGGGAGCGAAC..... 2013
446 etValLeuAsnLeuSerCysGluAlaSerGlnHisProArgProThrIle 462
      ::::::::::::::::::::
2014 .....CTCAACCTCTCTGCCACACCGGCGCTCAACCATCCCGCAGATAT 2058
463 SerTyrAsnValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArg 479
      ::::::::::::::::::::
2059 TCTTGGCGGTATCAATGGGATA.....CCGCGAGCA 2087
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuLur 495
      ::::::::::::::::::::
2088 ACACACACAAAGTCTCTTTATCGCCAAATACCGCAATATATACCGGA 2137
495 hrgLysValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
      ::::::::::::::::::::
2138 CC...TATGCTGTGTTGTCTCTACACTGGCTACTGGCGCGCAATATATCC 2184
511 IleuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerAs 527
      ::::::::::::::::::::
2185 ATAGTC..... 2190
527 nThrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAla 544
      ::::::::::::::::::::
2191 .AAGAGCATCACAGTCTCTGCATCTGGAACTTCTCTGCTCTCAGCTG 2239

```

```

544 snSerThrSerThrGluArgLysLeuProGluProGluSerArgGlyVal 560
      ::::::::::::::::::::
2240 GGGCGACT..... 2247
561 ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuGlyAl 577
      ::::::::::::::::::::
2248 .....GTCCGATCATGATGATGGAGTCTGCTGTT.....GGGCT 2279
577 aValLeu.TyrPheLeuTyrLysLysGly 586
      ::::::::::::::::::::
2280 TGCTGTGATATAGTTTATCTCGAGGGA 2308

seq_name: /cgn2_6/plodata/2/lna/5b.COMB.seq:US-08-458-356-145

seq_documentation_block:
; Sequence 145, Application US/08458356
; Patent No. 5942235
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,356
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 42506CURTMS
; INFORMATION FOR SEQ ID NO: 145:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-458-356-145

alignment_scores:
      Quality: 241.50      Length: 611
      Ratio: 0.918      Gaps: 23
Percent Similarity: 43.044      Percent Identity: 21.277

alignment_block:
US-09-653-961-2 x US-08-458-356-145 ..

Align seg 1/1 to: US-08-458-356-145 from: 1 to: 2349

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValThr 107
      ::::::::::::::::::::
751 AGGCTGAGCTGTCAATGCAATGCAACAGACCCCTCATCTATTCAATGTCAC 800

```

```

107 rProGlnAspGluArgIlePheLeuGlyGlnGlyLysArgProArgSer. 123
1081 AAGAAATGACACAGCAAGCTACAAATGTGAACCAAGCAACCCAGTAGTG 850
124 ....GInGluTyrArgIleGlnLeuArgValTyrLysAlaProGlnGlu 138
851 CCAAGGGCGCATGATTCAGTCATCTCAATGTCTCTATGCGCCGATGCC 900
139 ProAsnIleGlnValAsnProLeuGlyIleProValAsnSerLysGlu 155
901 CCCACCATTT.....TCCCTCTAAACACATCTTACAGATCAGGGGAAA 944
155 oGluGluValAlaIleThrCysValGlyArgAsnGlyTyrProIleProGln 172
945 TCTGAACCTCTCTCTGCCACGACCTCTAAC.....CCACCTGCCACAT 988
172 aIleTrrPrrLysAsnGlyArgProLeuLysGluLysAsnArgVal 188
989 ACTCTTGCTTGTCAATGGG.....ACT 1011
189 HisIleGlnSerSerGln.....ThrValGluLe 198
1012 TTCACACATTCACCCACAGAGCTCTTATCCCAACATCAGCTGTGATTA 1061
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
1062 TAGTGGATCTCTACCTGCCAAGCC..... 1086
215 LAspRysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
1087 .....CATACCTCA 1095
232 GlyAsnHisMetLysGluSerArgGluValIleThrValProValPheTyrPr 248
1096 GACACTGGCTCAATAGACACAGACAGATCAGATCAGATCTATGACGA 1145
248 oThrGluLysValIlePheGlnValGluProValGlyMetLeuLysGluG 265
1146 GCCACCCAAACCTTCATCACCAGCAACAATCCAAACCCGCGAGGATG 1195
265 LysAspArgValGluIleArgCysLeuAlaAspGlyAsnProProProHis 281
1196 AAGAGTGTCTACCTTAACCTCTGAACCTGAGATTCAGAACACAACTTAC 1245
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGlnGluThr 298
1246 CTGTGGTGGTAATATCATCAGAGCTCCCGGTCACTCCAGGCTCAGCT 1295
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGlnHis 315
1296 GTCCAAATGACACAGAGCCCTCAGCTCTACTCTCAGTCTCAAGAAGATGATG 1345
315 eArgIArgTyrGluCysGlnGlyLeuAspLysPheThrMetIleSerLeu 331
1346 TAGGACCTATGAGTGT...GGAATCCAGACGATTAAGTGTGAC... 1389
332 LeuSerGluProGlnGluLeuValAsnTyr...ValSerAspValAla 347
1390 CACACCGCACAGTCATCTGATGTCTCTATGCGCCAGACGACCCAC 1439
347 gValSerProAlaIleProGluArgGlnGlySerSerLeuThrLeuT 364
1440 CATTTCCCCCTCATACACCTATTACCGTCCAGGGGTGAACCTCAGCCTCT 1489
364 hrCysGluAlaGluSerSerGlnAspLysGluPheGlnIlePheArgGlu 380
1490 CCTGCAGTACGAGCCTTAACCAACCTGCACAGTATCTGGCTGATTTGAT 1539
381 GluThrGlnValLeuGlnValArgGlyProValLeuGlnLeuHisAspLe 397
1540 .....GGGAACATCCAGCAACACACAGAGCTTTATCTCCAAAT 1583

```

```

397 uLysArgGluAlaGlnGlyGlyTyrArgCysValAla..... 409
1584 CACTAGAGAAAGAACGGACTCTTATCTGCGAGCCAAATTAATCAGCCA 1633
409 ..... 409
1634 GNGGCCACAGCAGACTACAGTCAAGAACATCACAGTCTCTGCGGAGCTG 1683
409 ..... 409
1684 CCCAAGCCCTCATCTCCAGCAACAATCCAAACCCGTGAGAGCAAGA 1733
409 ..... 409
1734 TGCTGTGCTTACCTGTGAACCTGAGGCTCAGAACACAACTACTCTGT 1783
410 .....SerValPro..SerIleProGlyLeu..... 417
1784 GGTGGTAATGTGTCAGAGCTCCCAAGTCAAGTCCAGAGGCTGCAGCTGCC 1833
417 ..... 417
1834 AATGGCAACAGACCTCTACTCTATTCATGTCAAGAAATGACGCAAG 1883
418 .....AsnArgThrGlnLeuValAsnValAlaIlePhe 428
1884 AGCCTATGTATGTGAATCCGAACTCAGTGAAGTCAAAACCGCAGTACC 1933
429 GlyProProTrrPheAla..... 434
1934 CAGTACACCTGTGATGCTCTATGCGCGGACACCCCATCATTTCCCCC 1983
435 .....PheLysGluArgGlyValIleThrValLysGluAsnM 446
1984 CCAACATCTGTACTCTTGGGAGCGAAC..... 2013
446 eValLeuAsnLeuSerCysGlnAlaSerGlyHisProArgProThrIle 462
2014 .....CTCAACCTCTCTGCGACATGCGCTTAACCATCCCGCAGTAT 2058
463 SerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArg 479
2059 TCTTGGGTATCAATGGGATA.....CCGCAGCA 2087
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuGlnT 495
2088 ACACACACAAAGTTCTCTTATCGGCAAAATCAGCCCAATTAATACGGGA 2137
495 hrGlyValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
2138 CC...TATGCTGTGTTGTCTTAACCTTGCGTACTGCGCCAAATATTC 2184
511 lLeuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerHis 527
2185 ATAGTC..... 2190
527 nThrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAla 544
2191 AAGAGCATACAGATCTGTCATCTGAGACTGTCTCTGCTGTCTGACGCG 2239
544 snSerThrSerThrGluArgLysLeuProGluProGluSerArgGlyVal 560
2240 GGGCCACT..... 2247
561 ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuGluAl 577
2248 .....GTCCGATCATGATTTGAGATGCTGTGTT.....GGGT 2279
577 aValLeu.TyrPheLeuTyrLysLysGly 586
2280 TGCTGTGATATAGTTTATCTGACAGGA 2308

```

seq_name: /cgn2_6/prodata/2/ina/6B_COMB.seq: US-08-460-736-145

```

seq documentation block:
: Sequence 145: Application US/08460736
: Patent No. 6265189
:
: GENERAL INFORMATION:
: APPLICANT: Paolletti, Enzo
: APPLICANT: Tartaglia, James
: APPLICANT: Cox, William I.
: TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
: NUMBER OF SEQUENCES: 217
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Curtiss, Morris & Safford
: STREET: 530 Fifth Avenue
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10036
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,736
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/184,009
: FILING DATE: 19-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Frommer, William S.
: REGISTRATION NUMBER: 25,506
: REFERENCE/DOCKET NUMBER: 454310-2530
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 840-3333
: TELEFAX: (212) 840-0712
: TELEX: 425066CURTMS
: INFORMATION FOR SEQ ID NO: 145:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2349 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
:
: US-08-460-736-145
:
: alignment_scores:
:
:      Quality: 241.50      Length: 611
:      Ratio: 0.918      Gaps: 23
: Percent Similarity: 43.044      Percent Identity: 21.277
:
: alignment_block:
: US-09-653-961-2 x US-08-460-736-145 ..
:
: Align seg 1/1 to: US-08-460-736-145 from: 1 to: 2349
:
: 91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValrh 107
: |||||:::||||::: |||||::: |||||::: |||||::: |||||:::
: 751 AGGCGCGAGCTGTCACATGCGAAGGAGGAGCCCTCATCTTTCATGTCAC 8000
:
: 107 rProGlnAspGluArgGlyLeuPheLeuGlnGlyGlnGlyLysArgProArgSer 1233
: |:::|||||::: ||::: ||||::: ||||::: ||||::: ||||:::
: 801 AAGAATAATGACACAGCAAGCTACAAATGTGAAACCCGAACCCGATGAGTG 850
:
: 124 ....GlnGluTyrArgGlyGlnLeuArgValTyrIysAlaProGluGln 138
: ::::: ||::: ||||::: ||||::: ||||::: ||||:::
: 851 CCAGGCGCGAGTATTCAGTCATCATCTGAATGTCCTATATGCCCCGAGTGCC 9000
:
: 139 ProAsnIleGlnValAsnProLeuGlyIleProValAsnSerLysGluPur 155
: |||||::: |||||::: |||||::: |||||::: |||||:::
: 901 CCCACCATTT.....TCCCCTTAAACACATCTTACAGATCAGGCGGAAA 9444

```

155	ogluGluValAlaIaIaThCysValGluYrGsnGlnYrProIleProGlu	172
172	allIeTPTyYrLysAsnGluYrGProLeuYsgLulLysAsnArgVal	188
945	TCtGAACCTCTCCGCGCACGACGCTCTAAC.....CCACCTGCACACT	988
198	ACCTTGTTGTTGTCAAAGGG.....ACT	1011
189	HisIleGlnSerGln.....ThaValGluLeu	198
1012	TTTCAGCAAAATCCACCACCAAGAGCTTTATCCCAACATCATCTGTGAATA	1061
198	rSerGluYleuYrThrGlnSerIleLeuYsAlaGlnLeuValLysG	215
1062	TAGTGGATTCCTATACGTGCCAAGC.....	1086
215	LuaSplyAspAlaGlnPheTyTyrCysGluLeuAsnYrArgLeuProSer	231
1087CATAACTCA	1095
232	GlyAsnHisMetLysGlnSerArgGluValThrValProValPheTyT	248
1096	GACATGGCCCTCAATAGGACACACAGTACACGACGATCACTCATATGACGA	1145
248	oThrGluYsValTrrpleuGluValGluProValGlyMetLeuYsgLul	265
1146	GCCACCCAAACCTTCATACACACCAACAACTCCAAACCCCGTAGAGATG	1195
265	LYAspArgValGluIleArgGysLeuAlaSpGlyAsnProProHis	281
1196	AGGATGCTGTACCTTCACTGACCTGTGATTCGAAACACACACTAC	1245
282	PheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGlnGluThr	298
1246	CTGTGGTGGTAATAATACAGAGCCTCCGGTCACTCCACAGGCTGCAGCT	1295
298	rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysLHisS	315
1296	GTCCATATGCAACACAGGACCCCTCACTCATCTGATGCAACAGGATATG	1345
315	eArgIValTyrGluCysGlnGlyLeuAspLeuAspThrMetLieserLeu	331
1346	TAGGACCCCTATAGCT...GGAATCCAGAACGAATTAAAGTTGTTGAC...	1389
332	LeuSerGluProGlnGluLeuLeuValAsnTyT...ValSerAspValAr	347
1390	CACAGGACCCAGCTATCTGTGAATCTCTCATGAGCCCGACAGACCCAC	1439
347	GValSerProAlaAlaProGlnArgGlnGlnGlySerSerLeuThrLeuT	364
1440	CATTTCCTCCCTATACACTTATACCGTCCAGGGGTGAACCTCAGCCTCT	1489
364	hTrCysGluAlaLueSerSerGlnAspLeuGluPheGlnTrrpleuArgGlu	380
1490	CGTGCATGACGCTCTAACCCACCTGCACAGTATTTCTGGCTGATGAT	1539
381	GluThrGlyGlnValLeuGlnLurGlyProValLeuGlnHisAspLeu	397
1540GGGAACATCCAGGACACACACAGAGGCTTTATCTCCACAT	1583
397	uLysArgGluAlaGlyGlyTyTyrArgGysValAla.....	409
1584	CACGTAGGAAGAACAGCGGACCTGTATACCTGCCAGGCCAATAACTACGCA	1633
409	409
1634	GTGGCCACAGCAGAGACTACAGTCAAGACAAATCAGACTCTTGCGGAGCTG	1683
409	409
1684	CCCAAGCCCTCCATCTTCACAGCAACAACCTCCAAACCGTGAGGACAGGA	1733
409	409

```

1734 TGCTGGCCCTTACCTGTGAACTGAGGCTCAGAAACAACTACCTGT 1783
410 .....SerValPro.SerIleProGlu..... 417
1784 GGTGGTAATGTGACAGGCTCCAGTGTCCAGGCTGCAGCTGTCC 1833
417 ..... 417
1834 AATGGACAGACCTCACTATTCATATGTCAGAAATGAGCCAG 1883
418 .....AsnArgThrGlnLeuValAsnValAlaIlePhe 428
1884 AGCCTTGTATGTGAAATCAGAACTCACTAGTGCAGAACCCAGTACC 1933
429 GlyProProTrpMetAla..... 434
1934 CAGTACCCCTGGATGTCCTATGGCCGAGACACCCCATTCATTCGCC 1983
435 .....PheLysGluArgLysValTrpValLysGluAsnM 446
1984 CCAGACTCGCTTACCTTCCGGAGGAGAAC..... 2013
446 eValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 462
2014 .....CTCAACTCTCTCTCCACCTGCGCTTAACCCATCCCGCAGTAT 2058
463 SerTrpAsnValAsnGlyThrAlaSerGluAsnGlnAspProGlnArg 479
2059 TCTTGGCGCATCAATGGGATA.....CCGCAGCA 2087
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuGluTr 495
2088 AACACACACAGATTCTTTATTCGCCAATAATCAGCCCAATATTAACGGGA 2137
495 hGlyValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
2138 CC...TATGCCGTGTTGTGCTCTAATCTGGCTGACTGCGCCGCAATTAATCC 2184
511 IleLeuPheLeuGluLeuValAsnLeuThrThrProAspSerAs 527
2185 ATAGTC..... 2190
527 nThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAla 544
2191 .AAGAGCATCAGATCTCTGCTGAGACTTCTCTGCTGCTCAGAGTG 2239
544 snSerThrSerThrGluArgLysLeuProGluProGluSerArgGlyVal 560
2240 GGGCCACT..... 2247
561 ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuGlyAl 577
2248 .....GTCGCAATCATGATTGAGTGTCTGT.....GGGT 2279
577 aValLeu.TyrPheLeuTyrLysLysGly 586
2280 TGCTCTGATATAGTTTATCTCAGAGGA 2308
seq.name: /cgn2_6/protdata/2/1na/5b_comb.seq:us-08-602-725-35
seq_documentation_block:
; Sequence 35, Application US/08602725
; Patent No. 5965710
; GENERAL INFORMATION:
; APPLICANT: BODMER, WALTER F
; APPLICANT: DURBIN, HELGA
; APPLICANT: SNARY, DAVID
; APPLICANT: STEWART, LORNA MD
; APPLICANT: YOUNG, SUSAN
; APPLICANT: BATES, PAUL A
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES FOR USE IN
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF COLORECTAL CANCER

```

```

; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,725
; FILING DATE: 02-FEB-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB94/01816
; FILING DATE: 19-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9317423
; FILING DATE: 21-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36663
; REFERENCE/DOCKET NUMBER: 1090-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4091
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2097 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Human carcinoembryonic antigen
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2094
; US-08-602-725-35

alignment_scores:
Quality: 241.00 Length: 557
Ratio: 0.984 Gaps: 21
Percent Similarity: 43.986 Percent Identity: 22.083

alignment_block:
US-09-653-961-2 x US-08-602-725-35 ..
Align seg 1/1 to: US-08-602-725-35 from: 1 to: 2097

91 ArgLeuSerLeuGlnAspArgGlyAlaThrIleuAlaLeuThrGlnValTh 107
|||||:|||||: |||: |||||||: |||: |||: |||
556 AGCGTCAGCTGTCGAATGCAAGCAAGACCGCTCCTATGCGCCGAGTCC 605
107 rProGlnAspGluArgIlePheLeuGlnGlyIleProValAsnSerLysGluPr 123
|:|||||: |||: |||: |||: |||: |||
606 AAGAAATGACACAGCAAGCTACAAATGTGAAACCAAGAACCCAGTGAAGT 655
124 .....GlnGluTyrArgIleGlnLeuArgValTyrLysAlaProGluGlu 138
|||||: |||: |||: |||: |||: |||
656 CCAGGCGCAGTGAATCATCTGAAATGCTCTATGCGCCGAGTCC 705
139 ProAsnIleGlnValAsnProLeuGlnLysIleProValAsnSerLysGluPr 155
|||||: |||: |||: |||: |||: |||
706 CCCACCAT.....TCCCTCTAACAACATCTTACAGATCAGGGGAAA 749
155 oGluGluValAlaThrCysValGlyArgAsnGlyTyrProIleProGlnV 172

```

750 TGTGAACCTCTCTCCACGACGCTTAC.....CCACCTGCACGT 793
172 a111eTPTyrlYsAnGlYArGProLeuLYSGluGLYsAnArYAl 188
794 ACTCTTGTTGTCAATGGG.....ACT 816
189 H1s1eG1nSerSerGln.....ThrValG1uSe 198
817 TTCAGCAATTCACCCAAAGAGCTTTATCCCAACATCACTGTAATTA 866
198 rSerG1uLeuTYrThrLeuG1nSer1eLeuLYsAlaG1nLeuValLYSG 215
867 TAGTGATCTCTACGTGCACAGCC..... 891
215 lAsPlYsAlaG1nPhETyCYSGluLeuAsnTYrArGLeuProSer 231
892CATACCTCA 900
232 GlYAsnH1sMetLYSG1uSerArGluValThrValProValPhETyPr 248
901 GACACGCGCTCAATAGACACACAGTACGACGATCAGCTATACAGA 950
248 oThrGluLYsValTYrLeuGluValGluProValG1uMetLeuLYSGluG 265
951 GGCACCCAAACCTTCATCACCAGCAACAATCCAAACCCCTGGAGATG 1000
265 lYAsPArYAlG1uLeArGYsLeuAlaAsPGLYsAnProProh1s 281
1001 AGGATCTCTGACCTTACCTGTGAACCTGATTCAGAACACAACTAC 1050
282 PhEser1eSerLYSG1uAsnProSerThrArGluAlaG1uG1uTh 298
1051 CTGTGGTGGTAATATATACAGACCTCCGGTCACTCCAGCTGCAGT 1100
298 rThrAsnAsPAsnG1YValLeuValLeuG1uProAlaArGLYSGluH1s 315
1101 GTCCAAATGACAAACAGACCTCACTCACTCACTGACACAGAAATGATG 1150
315 eRGLYArGYrGLYCYSG1uG1nLYsLeuAsPLeuAsPhrMetLeuSer 331
1151 TAGAACCTCTATAGTGT.....GGATCCAGACGAATTAAGTGTGAC... 1194
332 LeuSerG1uProG1nLeuLeuValAsnTYr...ValSerAsPValAr 347
1195 CACAGGACCCAGTCACTCTGATGTCTCTATGCCCCAGACAGACCCAC 1244
347 gValSerProAlaAlaProG1uArG1nG1uLYsSerSerLeuThrLeuT 364
1245 CATTTCCCTCATACACCTATATCCGTCACGGGTTGACCTCAGCTCT 1294
364 hrcYSG1uAlaG1uSerSerG1uAsPLeuG1uPhE1nTrLeuArG1u 380
1295 CCGCAGATGAGCTCTAACCCACCTGCACAGATATCTTGCGTATGAT 1344
381 GluThrG1uValLeuG1uArG1YProValLeuG1nLeuH1sAsPLe 397
1345GGAAATCCAGCAACACACAAAGAGCTCTTATATCCAAAT 1388
397 uLYsArG1uAlaG1YGLYrArGYsValAla..... 409
1389 CACTGAGAAAGACAGGAGCTATACCTGCAGGCCAATTAACCTACCA 1438
409 409
1439 GTGGCCACACAGAGACTACAGTACAGACATCAAGTCTCGGAGCTG 1488
409 409
1489 CCCAAGCCCTCATCTCCAGCAACAACCTCCAAACCGTGAGAGACAAGA 1538
409 409

1539 TGTGTGGCCTTCACCTGTGAACCTGAGGCTCAGAAACAACTACTCT 1588
410SerValPro.Ser1eProG1YLeu..... 417
1589 GGTGGTAATATGTCAGAGGCTCCAGTCACTGCTCCAGGCTGCAGCTGTC 1638
417 417
1639 AATGGCAACAGAGACCTCACTATTCATATGTCACAAAGAAATGACCAAG 1688
418AsnArGThrG1nLeuValAsnValAla1ePhe 428
1689 AGCCTATGATATGGAATTCAGAACTCAGAGTGAAGTGAACCGAGGAC 1738
429 GLYProProTrpMetAla..... 434
1739 CAGTACCCCTGATGTCTCTATGGCGGACACCCCAATCATTTCCCC 1788
435PhelYSG1uArGLYsValTYrValLYSGluAsnM 446
1789 CCAGACTGCTCTTACTTTCGGGAGCGAAC..... 1818
446 eValLeuAsnLeuSerCYSGluAlaSerG1YH1sProArGProThr1e 462
1819CTCAACCTCTCTGCTCAGCTGCTCTAACCCATCCCGCAGAT 1863
463 SerTrpAsnValAsnGLYThrAlaSerG1uG1nAsPGLYsAnProG1nAr 479
1864 TCTTGGCTATCAATGGAT.....CCGACGA 1892
479 gValLeuSerThrLeuAsnValLeu...ValThrProG1nLeuLeuG1uT 495
1893 ACACACAAAGTCTCTTATCCGCAAAATCACCCAAATATATACCGGA 1942
495 hRGLYAlaG1uCYsThrAlaSerAsn...AsPLeuG1YLYsAnThrSer 510
1943 CC...TATGCTGTTTCTCTACTTGGCTACTGCGCGATATATTC 1989
511 lLeuPhELeuG1uLeuValAsnLeuThrThrLeuThrProAsSerAs 527
1990 ATA.....GTCAAGAGCATCAGACTCTCTCA...TCTGG 2021
527 nThrThrThrG1uLeuSer 533
2022 AACTTCTCTGCTCTCA 2040
seq_name: /cgr2_6/prodata/2/ina/5A_COMB.seq:US-08-389-459A-16
seq_documentation_block:
; Sequence 16, Application US/08389459A
; Patent No. 5817512
; GENERAL INFORMATION:
; APPLICANT: Morrow, Casey D. and Porter, Donna, C.
; TITLE OF INVENTION: ENCAPSIDATED RECOMBINANT POLIOVIRUS
; TITLE OF INVENTION: NUCLEIC ACID AND METHODS OF MAKING AND
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,459A
; FILING DATE: 15-FEB-1995
; CLASSIFICATION: 435


```
446 etValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 462
      |||||||
1927 .....CTCAACCTCTCTGCGCACTCGGCGCTTAACCAACCCCGAGTAT 1971
      |||||||
463 SerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspGlnArg 479
      |||||||
1972 TCTGGCGGTATCAATGGGATA.....CCGCAGCA 2000
      |||||||
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGluT 495
      |||||||
2001 ACACACACAGATCTCTTTATCCCAAAATACGCCCAATATATACGGGA 2050
      |||||||
495 hrgIyValGlyCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
      |||||||
2051 CC...TATGCCCTGTTGTCTTAAGTGGCTACTGCGCGCATATATTCG 2097
      |||||||
511 IleLeuPheLeuLeuValAsnLeuThrThrLeuThrProAspSerAs 527
      |||||||
2098 ATA.....GTCAAGACATACAGACTCTGCA...TCTGG 2129
      |||||||
527 nhrThrThrGlyLeuSer 533
      |||||||
2130 AACTCTCCGCTCTCTCA 2148

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-987-867A-16

seq_documentation_block:
; Sequence 16, Application US/08987867A
; Patent No. 6063384
; GENERAL INFORMATION:
; APPLICANT: C. Morrow et al.
; TITLE OF INVENTION: ENCAPSULATED RECOMBINANT VIRAL
; TITLE OF INVENTION: NUCLEIC ACID AND METHODS OF MAKING AND
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/987,867A
; FILING DATE: 09-DEC-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/087,009
; FILING DATE: 01-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: UAP-004CPDV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2203
; US-08-987-867A-16
```

```
alignment_scores:
      Quality: 241.00      Length: 557
      Ratio: 0.984      Gaps: 21
      Percent Similarity: 43.986      Percent Identity: 22.083

alignment_block:
US-09-653-961-2 x US-08-987-867A-16 ..

Align seg 1/1 to: US-08-987-867A-16 from: 1 to: 2220

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGluLeu 107
   |||||
664 AGCGTCGACGTCTGCAATGCAACAGACCCCTCACTCATTCATTAATGTCAC 713
   |||||
107 rProGlnAspGluArgGlyLeuPheLeuCysGlnGlyLysArgProArgSer 123
   |||||
714 AAGAAATGACACAGCAACGCTCAAAATGGAACCAACCAAGACCAAGTGA 763
   |||||
124 ....GlnGluThrArgGlyLeuGlnLeuArgValTYrLysAlaProGluGlu 138
   |||||
764 CCAGCGCAGATGATTCAGTCAATCCGAAATGTCCTATGCGCCGGAATGCC 813
   |||||
139 ProAsnIleGlnValAsnProLeuGlyTLeProValAsnSerLysGluPr 155
   |||||
814 CCACACAT.....TCCCTCTAAACACATCTTACAGATCAGGAGGAAAA 857
   |||||
155 oGluGluValAlaThrCysValGlyArgAsnGlyTYrProIleProGlu 172
   |||||
858 TCTGAACCTCTCTGCCATGCAGCCTCTAAC.....CCACCTGCACAGT 901
   |||||
172 alleTrpTYrLysAsnGlyArgProLeuLysGluGluLysAsnArgVal 188
   |||||
902 ACTCTGCTTGTCTCAATGG.....ACT 924
   |||||
189 HisIleGlnSerSerGln.....ThrValGluSer 198
   |||||
925 TTCACGCAATCCACCAAGAGCTCTTATCCCAACATCACTGGAATTA 974
   |||||
198 rSerGlyLeuTYrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
   |||||
975 TAGTGATCTCTATACGTGCCAAGCC..... 999
   |||||
215 LuAspLysAspAlaGlnPheTYrCysGluLeuAsnTYrArgLeuProSer 231
   |||||
1000 .....CATACTCA 1008
   |||||
232 GlnAsnHisMetLysGluSerArgGluValThrValProValPheTYrPr 248
   |||||
1009 GACACTGGCTCATATAGACACAGTCACGACGATCAGCTATATGACAGA 1058
   |||||
248 cThrGlnLysValTrpLeuGluValGluProValGlyMetLeuLysGlu 265
   |||||
1059 GCCACCCCAACCTTCATCACCAAGCAACACTCCAACTCCCGGTGGAGGATG 1108
   |||||
265 LysAspArgValGluIleArgCysLeuAlaAspGlyAsnProProGlnHis 281
   |||||
1109 AGAATGCGTGAACCTTAACCTGATACCTGATGATTCAGAACACAACTTAC 1158
   |||||
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluGlu 298
   |||||
1159 CTGTGGTGGGTAAATATCAAGAGCTCCCGGTACGTCCAGGCTGACGT 1208
   |||||
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
   |||||
1209 GTCCAAATGACACAGACCCCTCACTCACTCACTCACTCACTCACTCACT 1258
   |||||
315 ergIyArgTYrGlyCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
   |||||
1259 TAGGACCTTATGAGTGT...GGAATCCAGAACGAAATTAAGTGTGAC... 1302
   |||||
332 LeuSerGluProGlnGluLeuValAsnTYr...ValSerAspValArg 347
   |||||
```

```

1303 CACAGGACCCAGTCATCTCGAATTCCTCTATGAGCCAGACGCCAC 1352
347 gvalserProAlaIalProGluAlaGlnGluGlySerSerLeuThrLeu 364
1353 CATTTCCTCCCTATACACCTATTACCGTCAGGGGTGAACCTCAGCTCT 1402
364 hrcysglnuIaGluSerSerGlnAspLeuGluIupheGlnIupheAlaGlu 380
1403 CCGGCATGACAGCCTGTAAACCAACCTGGACAGATATCTTGCGTATGAT 1452
381 GluThrGlyValLeuLeuGluArgGlyGlyProValLeuGlnLeuHisAspLe 397
1453 .....GGGAAATCATCAGCAACACACACAAAGAGCTTTATCTCAACAT 1496
397 uLyAsrArgIuAlaGluGlyGlyTyrArgCysValAla..... 409
1497 CACTGAGAAAGAACAGCGGACTATATACCTGCCAGGCCAATACATCAGCA 1546
409 ..... 409
1547 GTGGCCACAGACGACTACGTCAAGACAAATCAGATCTCTCGGAGCTG 1596
409 ..... 409
1597 CCCAAGCCTCATCTCCAGCAACAATCTCAAAACCCGTGGAGACAAAGA 1646
409 ..... 409
1647 TGCTGTGGCCTTACCTGTGAACCTGAGGCTCAGGCTCAGAACAACTAACGT 1696
410 .....SerValPro.SerLeuProGlyLeu..... 417
1697 GGTGGGTAAATGGTGCAGAGCCTCCAGTCAAGTCCAGGCTGACAGCTGCC 1746
417 ..... 417
1747 AATGGCAACAGACGACCTCATCTGTATTCATGTGCACAAAGAAATGAGCCAA 1796
418 .....AsnArgThrGlnLeuValAsnValAlaIalLeuPhe 428
1797 AGCCTATGTATGTGAATTCACAAACACTGACAGTGCAAACCGCAGTGACC 1846
429 GlyProProThrPheAla..... 434
1847 CAGTCACCTGGATGTCCTCTATGGGCGCGAGACACCCCATCTTTCCCC 1896
435 .....PheLysGluArgIysValIupheValIupheLysGluAsnM 446
1897 CCAAGACTCGCTTACCTTTCGGGAGAGAAC..... 1926
446 etValLeuAsnLeuSerGysGluAlaSerGlnHisProArgProThrIle 462
1927 .....STCAACSTCTCTGCGACATGGCGCTCTAAACCCATCCCGCAGAT 1971
463 SerTyrAsnValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArg 479
1972 TCTGTGGGTATCAATGAGATA.....CCGCACGA 2000
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGlyT 495
2001 ACACACACAAAGTCTCTTTATCCCAAAATTCACGCCAATAATTAACGGGA 2050
495 hrcGlyValGlyCysThrAlaSerAsp...AspLeuGlyLysAsnThrSer 510
2051 CC...TATGCGCTGTGTTGCTCAACTTGCGGTACAGCGCGGCATATATTC 2097
511 IleLeuPheLeuGluLeuValAsnLeuThrThrThrLeuThrProAspSerAs 527
2098 ATA.....GTCAAGAGCATACACAGTCTGTGCA...TCTGG 2129
527 nThrThrThrGlyLeuSer 533
2130 AACTTCTCTGTGGTCTCA 2148

```


Align seg 1/1 to: US-08-468-856B-5 from: 1 to: 2839

```
91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValThr 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 AGGCTGACGCTGCTCCATGAGCAGACGACCCCTCATCTTATTCATGATGAC 621
107 rProGlnAspGluArgGlyLeuPheLeuCysGlnGlyLysArgProAlaSer. 123
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
622 AAGAATGAACAAGCAAGCTCAATGTAAGTAAGCAAGCAAGCAAGTAAG 671
124 .....GlnGluThrArgGlyLeuGlnLeuValThrLysAlaProGluGlu 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
672 CAGGGGCAAGTATTCATGATGCTGATGATGCTGATGCTGATGCTGATGCT 721
139 ProAsnGlnGlnValAsnProLeuGlyLysProValAsnSerLysGluPro 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
722 CCGACCAAT.....TCCCTGTAAACACATCTTACAGATCAGGGGAA 765
155 OGluGluValAlaThrCysValGlyArgAsnGlyLysProLysProGlu 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
766 TGTGAACCTCTCTCCGACGACGACCTTAC.....CCACCTGACAGT 809
172 AlLeuThrLysAsnGlyLysArgProLeuLysGluLysAsnArgVal 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
810 ACTCTGTGTGTCAATGGG.....ACT 832
189 HisIleGlnSerGln.....ThrValGluSer 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
833 TTCCACGACATCCACCCAGAGCTCTTATCCACGACATCAGTGAATTA 882
198 rSerGlyLeuThrLysGlnSerIleLeuLysAlaGlnLeuValLysG 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
883 TAGTGATCTTACGTGCGCAAGCC..... 907
215 LysAspLysAlaGlnPheThrCysGluLeuAsnThrArgLeuProSer 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
908 .....CATACCTCA 916
232 GlyAsnHisMetLysGlnSerArgGluValThrValProValPheThr 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
917 GACACTGCTCATATGAGCAGACAGTCCAGTCAAGTCTATGTCAGA 966
248 ThrGlyLysValThrLeuGluValGluProValGlyMetLeuLysGlu 265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
967 GCCACCCAAACCTTCATCACCAGACACCTCCAAACCCGTTGGAGGATG 1016
265 LysAspArgValGluLeuArgCysLeuAlaAspGlyAsnProProPheHis 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1017 AGGATGCTGAGCTTAACTGTAACCTGTGACCTGATTCAGAAACAACCTAC 1066
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGlnGluGluThr 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1067 CTGTGGTGGTAATATATATACAGGCTCCCGGTGATGCCAGGTGAGCT 1116
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaGlyLysGluHis 315
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1117 GTCCAAATGACAGACGACCTCTACTCTAGTGTACAGGAAGATGATG 1166
315 ArgLysArgGlyLysGlnGlyLysLeuAspThrMetIleSerLeu 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1167 TAGAGCTATAGAGTGT...GGAATCCAGAAAGCAATTAACTGTGAC... 1210
332 LeuSerGluProGlnGluLeuLeuValAsnThr...ValSerAspValArg 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1211 CAGAGGAGACCATGCTCATCTGATGCTGATGCTGATGCTGATGCTGATG 1260
347 gValSerProAlaAlaProGluArgGlnGlyLysSerSerLeuThrLeu 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1261 CATTTCCCTCATACACCTATTAACCTGCGGAGGAGTAACTGAGCTCT 1310
364 hrCysGluAlaGlnSerSerGlnAspLeuGluProGlnThrLeuValGlu 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1311 CTGGCATGAGACCTCTTAACCACTGACAGATATCTTGCTGATGAT 1360
```

```
381 GluThrGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLe 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1361 .....GGGACATCCAGACACACACAGAGAGCTCTTATCTCCACAT 1404
397 uLysArgGluAlaGlyGlyGlyLysArgCysValAla..... 409
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1405 CACTGAGAAAGACAGGAGCTATACCTGCGGACCAATATCAGCA 1454
409 ..... 409
1455 GTGGCCACAGCAGGACTACAGTCAAGCAATCAAGTCTCTGCGGAGCTG 1504
409 ..... 409
1505 CCCAAGCCTCATCTCCAGACACACTCAACCCGTTGGAGGACAGGA 1554
409 ..... 409
1555 TGCTGTGACCTTCACTGTGAACTGAGGCTCAGAAACACACTACCTGT 1604
410 .....SerValPro.SerIleProGlyLeu..... 417
1605 GTGTGGTAATGTGACAGCTCCCACTGATCCAGCTGAGCTGTCC 1654
417 ..... 417
1655 AATGCAACAGGACCCCTCATCTATCAATGTCAACAAATGACCGAAG 1704
418 .....AsnArgThrGluLeuValAsnValAlaIlePhe 428
1705 AGCTTATGATGTGAAATCCAGAACTGATGATGCAACCCGAGTGGCC 1754
429 GlyProProThrMetAla..... 434
1755 CAGTCACTGATGTCTCTATGCGCGGACACCCCATCATTTCCCC 1804
435 .....PheLysGluArgLysValThrValLysGluAsnM 446
1805 CAGAGCTGCTTACCTTTCGGGAGCAGAC..... 1834
446 eValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 462
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1835 .....CTCAACCTCTCCGACCTGCGCTCTTAACTCCCGGAGTAT 1879
463 SerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArg 479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1880 TCTTGGCTATCAATGGATA.....CCGACAGA 1908
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuGluThr 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1909 ACACACACAAAGTCTCTTATCGGCAAAATCAGCCCAATATATACGGA 1958
495 hrGlyValGlnCysThrAlaSerAsn...AspLeuGlyLysAspThrSer 510
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1959 CC...TATGCTGTGTTGCTCTTACTTGGCTACTGCGCCGCAATATTC 2005
511 IleLeuPheLeuGlnLeuValAsnLeuThrThrLeuThrProAspSerAs 527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2006 ATA.....GTCAAGAGCATCAGCTCTGCA...TCGG 2037
527 nThrThrThrGlyLeuSer 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2038 AACTCTCTCTGCTCTCA 2056
```

seq_name: /cgn2_6/ptodata/2/ina/fa_COMB.seq:us-08-468-859A-5
seq_documentation_block:
Sequence 5, Application us/0846859A
Patent No. 6022958
GENERAL INFORMATION:
APPLICANT: Barnett, Thomas; Eiting, James; Komarck, Michael;
Kreischer, Axel


```
409 ..... 409
1455 GTGGCCACGACGACTACAGACATCAAGTCTGTGGGACGTG 1504
409 ..... 409
1505 CCCAAGCCCTCCATCTCCAGCAACATCCAAACCCCTGAGACAGAA 1554
409 ..... 409
1555 TGTCTGTGGCTTACCTGTGAACCTGAGACCTCAACACACCTACTCT 1604
410 ..... 417
1605 GGTGGTAAATGCTACAGACCTCCAGACGTCCAGGCTGACGTGCTCC 1654
417 ..... 417
1655 AATGGCAACGACCCCTCACTCTATTCAATGTCAAGAAATGACGCAAG 1704
418 ..... 428
1705 AGCCTATGTATGTGAATCCAGAACTCAGTGAAGCAACGACGAGTACC 1754
429 GlyProPorphyrin..... 434
1755 CAGTACACCTGATGTCCTCTATNGGCGGACACCCCATCATTTCCCCC 1804
435 ..... 446
1805 CCAGACCTGCTTACTCTGCGGACGAC..... 1834
446 etValLeuSerGluValSerGluValSerGluValSerGluValSer 462
1835 ..... 1879
463 SerTrpValSerGluValSerGluValSerGluValSerGluValSer 479
1880 TCTTGGCGTATCAATGGATA..... 1908
479 gValLeuSerThrLeuValSerGluValSerGluValSerGluValSer 495
1909 ACACACGACGCTCTTATGCGCAAAATCAGCGCAAAATATATACGGGA 1958
495 hrGlyValGluCysThrAlaSerAsn...AspLeuGlyIleValSer 510
1959 CC...TATGCTGTGTTGCTCTTACTGCTGCTGCTGCTGCTGCTGCTG 2005
511 IleLeuPheLeuGluValAsnLeuThrThrLeuThrProAspSerAs 527
2006 ATA..... 2037
527 nrThrThrGlyLeuSer 533
2038 AACTCTCTGCTCTCA 2056

seq_name: /cgn2_6/prodata/2/ina/PC/US.COMB.seq: PCT-US93-00031-14

seq_documentation_block:
Sequence 14 Application PC/TUS9300031
GENERAL INFORMATION:
APPLICANT: Osborn, Laurelee
APPLICANT: Benjamin, Christopher D.
TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
NUMBER OF SEQUENCES: 23
IMMUNOGLOBULIN-LINE DOMAIN OF VCAM1
CORRESPONDENCE ADDRESS:
ADDRESS: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00031
FILING DATE: 19930112
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 715-1000
TELEFAX: (312) 715-1234
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2208 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2208
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 73..2208
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1..72
PCT-US93-00031-14

alignment_scores:
Quality: 237.50 Length: 704
Ratio: 0.745 Gaps: 30
Percent Similarity: 45.312 Percent Identity: 20.739

alignment_block:
US-09-653-961-2 x PCT-US93-00031-14 ..
Align seq 1/1 to: PCT-US93-00031-14 from: 1 to: 2208

33 ProGluLeuValGluVal.....GluValGlySe 42
||||| ||||| ||||| |||||
334 CCAGAACGGGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 383
42 rThrAlaLeuLeuLysCysGlyLeuSerGlnSerGlnSerGlnSerGlnSer 59
||||| ||||| ||||| |||||
384 GAACCTTACCTACGCTGC.....CAGGTGGAGGTGGGCGACCCC 424
59 IsValAspTrpPheSerValHisGlyGluValSerGlnSerGlnSerGln 75
||||| ||||| ||||| |||||
425 GGGCGAAC.....CTCACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 453
76 ValArgGlnGlyGlnGlyGlnSerGluProGlyGluTyrGluGlnArg 92
||||| ||||| ||||| |||||
454 GGGAGAGAGCTCATGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 503
92 uSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValThrProG 109
||||| ||||| ||||| |||||
504 CTCCTCGAAGAACCAAGATTTGGAAGTACCTTACTCTGCTGATTCAGG 553
109 LnsAspGluArgIlePheLeuGlyGlnGlyLysArg..... 120
554 ATATTGGAAAGTCTGTTGTCGAGGTAAATTACCACTTATGATAATG 603
121 .....ProArgSerGlnGluTyrArgIleGlnLeuValTyrGly 134
604 GATTCTGTGCCACAGTAAAGGAGGCTGTAAAGAAATGCAAGTCAAT 653
134 sAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 151
```



```

68 uLySArGThrLeuIlePheArGValArgInGInGInGInSerGluP 85
|||||
203 GAAGGGAGGCTCATCTTCCGTGTGGCCAGGGCCAGGGCCAGACGAAAC 252
85 rGIGLyIuYrGInGInArGLeuSerLeuGInAsPArGIGLyAlaThrLeu 101
|||||
253 CTGGGGAGTACGAGGAGGGGCTCAGCTCTCAGACAGAGAGGGCTTACTCTG 302
102 AAlaLeuThrGInValIThrProGInAsPGLuArgIlePheLeuCyGInGIn 118
|||||
303 GCCCTGACTCAAGTCAACCCGCCCAAGCGAGCCATCTTCTGTGCCAGGG 352
118 yLySArGPArGArGSerGInGInuYrArGIleGInLeuArGValIYrLySA 135
|||||
353 CAAGGCCCTCGGTCCCAAGAGTACCGCATCCAGCTCCGCGCTTCAAGAA 402
135 lArProGInuYrProAsnIleGInValAsnProLeuGInYIleProValAsn 151
|||||
403 CTCGGGAGGAGCA .AACATCCAGGTCAACCCCTGGGCGATCCCTGTGAAC 451
152 SerLySGInuYrProGInuYrAlaThrCySAIGLyArGAsnGInYrPr 168
|||||
452 AGTAAAGAGCTGAGAGGCTCGTACCTGTGTAGGAGGAGAACGGGTACCC 501
168 oIleProGInValIleTrPTrLySAAsnGInYrArProLeuLySGInuYr 185
|||||
502 CATCTCTCAAGTCATCTGTGTACAGAAATGCCGCGCTCGAAGAGAGAA 551
185 ySAAsnArGValHisIleGInSerSerGInThrValGInuSerGInLeu 201
|||||
552 AGAACGGGTCCACATTCAGTCGTCCCAAGACTGTGAGTGCAGATGGTTTG 601
202 TyrThrLeuGInSerIleLeuLySAInGInuYrValLySGInAsPlySA 218
|||||
602 TACACCTTCAGAGTATCTTGAAGGCACAGCTGTGTAAGAAAGACAAAGA 651
218 pAlaGInPheTyCYSGInuYrLeuAsnTyArGLeuProSerGInAsnHis 235
|||||
652 TGGCCAGTTTACTGTGAGCTCAACTACAGCGGCTGCCCATGGGAACCA 701
235 eY .LySGInuSerArGInuYrAlaThrValProValPheTyrProThrGInuYr 251
|||||
702 TGAAGAGAGTCCAGGAGAGTCAACCGCTCTTCTTACCACCAAGAAAA 751
251 sValITrPLeuGInuYrAluProValGInuYrMetLeuLySGInuYrAsPArG 268
|||||
752 AGTGTGGCTGGAAGTGAAGCCCGTGGAAATGCTGAAGAGAGGAGACCGG 801
268 aIGInuYrLeuArGLeuAlaAsPGLyAsnProProProHisPheSerIle 284
|||||
802 TGGAAATCAAGGTGTGGCTGATGAGCAGCCCTCCACACACTTCACACATC 851
285 SerLySGInAsnProSerThrArGInuYrAlaGInuYrGInThrThrAsn 301
|||||
852 AGCAACCAAGAACCCAGACACAGGAGGAGAGAGAGACAAAC .AACAA 900
301 pAsnGInuYrAlaLeuValLeuGInuYr .ProAlaArGInuYrHisSerGInuYr 317
|||||
901 CAACGGGGTCTCTGTCTGAGCTGCCCGGAAGAACACATGGGSG . . . 946
317 rGTTrGInuYrGInGInLyLeuAsPLeuAsPThrMetIleSerLeuLeuSer 333
|||||
947 . . .TATGAAATGTCAAGSGCTGG . . . . .ACTGAAACATGATATYCTST 985
334 GluProGInuYrLeuLeuValAsnTyrVal .SerAsPValArGValSerP 350
|||||
986 GAGTGAACACAGGAAATGTGTGACTATGTTT . . . . .GTGGATK 1026
350 roAlaAlaProGInuYrGInGInuYrSerSerLeuThrLeuThrCySGInu 366
|||||
1027 ATCCGGCGCCTGAGAAAGAGACACAGCTCAGCTGACTGTGAGCAAAATA 1076

```

```

367 AlaGInuSerGIn 371
|||||
1077 GCAGACTCGATTTCG 1091

seq_name: gb_est1:AL556415
seq_documentation_block:
LOCUS AL556415 956 bp mRNA linear EST 16-FEB-2001
DEFINITION AL556415 LRT_NFL006.PL2 Homo sapiens cDNA clone CS0DK004YG06.5
prime, mRNA sequence.
ACCESSION AL556415
VERSION AL556415.1 GI:12899069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 956)
AUTHORS Li,W.B., Gruber,C., Jesse,J., and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..956.
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK004YG06"
/tissue_lib="LRT_NFL006.PL2"
/feature_type="placenta"
/note="Vector: PCWVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 214 a 272 c 292 g 171 t 7 others
ORIGIN

alignment_scores:
Quality: 1556.00 Length: 320
Ratio: 4.928 Gaps: 0
Percent Similarity: 98.750 Percent Identity: 97.812

alignment_block:
US-09-653-961-2 x AL556415 ..

Align seg 1/1 to: AL556415 from: 1 to: 956

1 MetGlyLeuProArGLeuValCysAlaPheLeuLeuAlaIacCyCy 17
|||||
1 ATGGGGCTTCCAGAGCTGTGCTGGCGCTTCTGCTGCGCGCTCTCTCTG 50
17 sCySPArGValAlaGInuYrAlaGInuYrAlaGInuYrProAlaProG 34
|||||
51 CTGTCTCGCGTGGGGGTGTGGCGGAGAGCTGAGAGCGCTCGGCTG 100
34 lUeUValGInuYrAluGInuYrSerThrAlaLeuLeuLyCySGInuYr 50
|||||
101 AKCTGTGTGAGGTGAATKGGACAGACAGCCCTTGAATKGGGCTC 150
51 SerGInSerGInGInLyAsnLeuSerHisValAsPTrPheSerValHisLy 67
|||||
151 TCCAGTCCCAAGCAACCTCAGCCATGTGAGCTGTCTTCTGTTCACAA 200
67 sGluYsArGThrLeuIlePheArGValArgInGInGInGInSerG 84

```



```

201 GAGAGAGGAGGAGCGTCACTTCCTGCTGCGCCAGGCGCCAGGCGCCAGGCG 250
84 LUPROGLUTYRGUGLInArgLenseLInAspArgLysAlaThr 100
251 AACCTGGAGACGACGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgLysLeuGly 117
301 CTGGCCCTGACTCAAGTACACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 350
117 nglYlsArGProArGserGInglutYrArGILegInLeuAqValTYRL 134
351 GGGCAAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400
134 yAlaLProGluLuproAsnLleGInValAspProLeuGlyLeuProVal 150
401 AAGCTCCGAGAGM .GCAGACATCCAGTCAACCCCTGGGCTCCCTGTX 449
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTY 167
450 AACAGTAAAGGAGACCTGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 499
167 rProLleProGlnValLleTPYrLYsAsnGlyArqProLeuLysGlu 184
500 CCCCATTCCTCAAGTCAATCTGTGTCAAGATGGCCGCTGTAAGAGAGAG 549
184 LuLysAsnArqValHsTleGInSerSerGInThrValGluSerSerGly 200
550 AGAAGACCGGGGTCCACATTCAGTCCGACGACGCTGTGAGTGCAGTGT 599
201 LeuTYrThrLeuGlnSerLleLeuLysAlaGlnLeuValLysGlnAsp 217
600 TTGTACACCTTGACAGATTCATTCAGAGCACAGCTGTGTTAAAGAGACAA 649
217 sAspAlaGlnPheTYrCysGluLeuAsnTYrArGLeuProSerGlyAsn 234
650 AGATGCCCACTTACTGTGAGTCACTACCGCTGCTGCTGCTGCTGCTGCT 699
234 IsMetLysGlnSerArqGluValThrValProValPheTYrProThrGlu 250
700 ACATGAGAGAGTCCAGGAGATCCAGCTGCTGCTGCTGCTGCTGCTGCTG 749
251 LysValThrLeuGlnValGluProValGlyMetLeuLysGlnLysAsp 267
750 AAAGTGTGCTGAGAGTGAAGCCCTGCGAATCTGAGAGAGAGAGAGAGAG 799
267 gValGlnLeuArgCysLeuAlaAspGlyAsnProProPheHisSer 284
800 CGTGGAATCAAGTGTGTGCTGATGCAACCTCCACCACTTCACCA 849
284 LeSerLysGlnAsnProSerThrArgLysAlaGlnGlnGlnGlnGln 300
850 TCARCAAGCAGAACCCAGC .ACCAGGAGGAGAGAGAGAGAGAGAGAAC 897
301 AspAsnGlyValLeuValLeuGlnProAlaArgLysGlnHisSerGly 317
898 GACAAAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946
317 gTYrGlnCys 320
947 CTATGATATCT 956
seq_name: gb_est1:AL541635
seq_documentation_block:
LOCUS AL541635 948 bp mRNA linear EST 16-FEB-2001
DEFINITION AL541635 LTI-FL002_P11 Homo sapiens cDNA clone CS0DE003YA16 5 prime
, mRNA sequence.
ACCESSION AL541635
VERSION AL541635.1 GI:12872890
KEYWORDS EST.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 948)
AUTHORS L.I.W.B., Gruber,C., Jesssee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..948
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE003YA16"
/location="LTI-FL002_P11"
/lab_host="DH10B"
/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end and enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetechn.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 213 a 271 c 296 g 166 t 2 others
ORIGIN
alignment_scores:
Quality: 1528.00 Length: 314
Ratio: 4.945 Gaps: 2
Percent Similarity: 98.408 Percent Identity: 98.408
alignment_block:
US-09-653-961-2 x AL541635
Align seg 1/1 to: AL541635 from: 1 to: 948
1 MetGlyLeuProArGLeuValCysAlaPheLeuLeuAlaIcysCysG 17
|||||
12 ATGGGGCTTCCAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61
|||||
17 sCysProArqValAlaGlyValProGlyGlnAlaGlnGlnProAlaProG 34
|||||
62 CTGTCTCGCGGTGCGGGGTGCGGAGGCTGAGAGGCTGAGGCTGCTG 111
|||||
34 LuLeuValGlnValGlnValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
112 ACCTGTGTGAGGTGAGAGTGGGAGCACAGCCCTTGTGAGAGTGGGCTG 161
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHis 67
|||||
162 TCCCACTCCCAAGGACACCTCAGCATGTGATGCTTTCCTGCTGCTG 211
|||||
67 sGlnLysArqThrLeuLlePheArqValArgGlnGlnGlnGlnGlnGln 84
|||||
212 GGAAGAGCGGAGCGTCATCTTCCTGTGCGCCAGGCGCAGGCGCAGAG 261
|||||
84 LuProGlyLutYrGlnGlnArgLeuSerLeuGlnAspArgLysAlaThr 100
|||||
262 AACCTGGAGATCAAGACGAGGCTGAGCTCCAGAGACAGAGAGGCTACT 311
|||||
312 CTGGCCCTGACTCAAGTCAACCCCAAGAGAGAGAGGATTTCTTGTGCA 361
|||||
117 nglYlsArGProArGserGInglutYrArGILegInLeuAqValTYRL 134
|||||
362 GGGCAAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
|||||

```

134 ysalaprogliupproasniieglnvalasproliengliyleproval 150
 412 AAGTCCGGAGAGANCA.AACATCCAGGTCAACCCCTGGGCAATCCCTGTG 460
 151 Asnerlysgliuprogliuvalalathrcysvalglarsanglty 167
 461 AACGTAAGAGAGCTGAGAGGTGCTACCTGCTGTAAGAGAGAGGTA 510
 167 rprolietroglnvalietptryllysanglyarproleuylsglug 184
 511 CCCCATCCCTCAAGTATCTGCTACCAAGATGGCCCGCTCGAAGAGAG 560
 184 lulyasnarargvalhisileginserserghnthrvalglusersgly 200
 561 AGAAGAACCGGTCACATTCAGTGTGTCAGAGAGCTGGAGTGCAGTGT 610
 201 leuylthrlenglnserileuylvalaglnleuvalylsglualsply 217
 611 TTGTACACCTTCGACAGTATCTCGAAGCAGCAGCTGTTAAAGAGAGCA 660
 217 sasplalaglnphetyrcysgluleuasnyrargleuproserylsan 234
 661 AGATGCCAGTATTACTGTAGCTCAACTACCGGCTGCCAGTGGAGACC 710
 711 ACATGAAGAGAGTCCAGGAGAGTCAACGCTCCTGTTTCTACCCAGCA 760
 250 ulysvaltrpleuglnvalgluprovalglymetleuylsglualspla 267
 761 AAAAGTGGCTGGAGAGTGGAGCCGCTGGAGATGCTGAAGAGAGGAGC 809
 267 rglvalglulearpcysleualaspglyasnprowprowprowprowprow 283
 810 C.GTGGAAATCAAGGTGTTGGCTGATGGCAA.CCTCACACACACTTCAC 858
 284 lleserlysglnasnprowprowprowprowprowprowprowprowprow 300
 859 ATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 908
 300 naspsan.glyvalleuvalleuglnproalaarglys 312
 909 CGACACAGGCTCTCTGCTGTGAGAGCTGCGGAGAG 946
 seq_name: gb_est1:AL554839
 seq_documentation_block: 1022 bp mRNA linear EST 16-FEB-2001
 LOCUS AL554839 LTI_NFL006.PL2 Homo sapiens cDNA CS001087Yg10 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL554839
 VERSION AL554839.1 GI:12896003
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1022)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1..1022
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS001087Yg10"
 /clone_lib="LTI_NFL006.PL2"
 /tissue_type="Placenta"

/note="Vector: pCMVSPORT 6; Site:1: NotI; 1st strand cDNA
 was primed with a NotI oligo(dT) primer. Five prime end
 enriched double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifestech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 237 a 306 c 289 g 184 t 6 others
 ORIGIN

alignment_scores:
 Quality: 1520.00 Length: 297
 Ratio: 5.118 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.327

alignment_block:
 US-09-653-961-2 x AL554839 ..
 Align seg 1/1 to: AL554839 from: 1 to: 1022

350 PROALAAPROGLIUPROGLIUGLIYSERSETHRLTHRCYSGI 366.
 2 CCCGACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 51
 366 ualagluserserghnthrvalgluphleglnthrcysglulthrg 383
 52 GCGAG 101
 383 lylglnvalleuglnuarglyprovalleuglnleuvalaspluysary 399
 102 GCCAGGTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 151
 400 GUALAGLYGLYTRYRARGYSVALASERVALPROSERLIEPROGI 416
 152 GAGGACAG 201
 416 yleuasnarargthrglnleuvalasvalalalepheyglyproptrpm 433
 202 CCTGACACAG 251
 433 etalaphelysglnuarglyvaltrpvalylsglnuasmetyalleuasn 449
 252 TGGCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 301
 450 leusercysglualaserglylasprowprowprowprowprowprowprow 466
 302 CTGTCTTGTGAAGCGTCAGGGCAGCCCGGCGCACCATCTCTGTGAACGT 351
 466 laseglythralaserglnaspglnaspprowprowprowprowprowprow 483
 352 CAGGAG 401
 483 hrleuasnvalleuvalthrcysgluleuvalthrcylvalglucys 499
 402 CCCGATATGCTCTGTGACCCCGAGAGCTGTGGAGAGAGAGAGAGAGATG 451
 500 ThrAlasernasplleuglyllysasnThrserileuvalleuglule 516
 452 ACGGCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501
 516 uvalasleuvalthrcysgluleuvalthrcysgluleuvalthrcysgl 533
 502 GGTCAATTTACACCTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 551
 533 erthrcysgluleuvalthrcysgluleuvalthrcysgluleuvalthrc 549
 552 GCATCTTCAG 601
 550 Arglylsleuprogliuprogliuprogliuprogliuprogliuprogli 566

```
|||||
602 AGAAGCTCCGAGCCGAGAGCCGGCGTGTGATGTGCGTGTAT 651
566 eValcyslleuValleuAlaValleuGlyAlaValleuTyrrpheleut 583
652 TGtGTGATCTGCTGCTGCGGCTGCTGCGCTCTCTCTCTCTCTCT 701
583 ytlslYsglYlYslleuProCysArgArgSerGlyLysGlnGluThr 599
702 ATAAAGAGGCAAGCTGCGCTGCGGCGCTCAGGAGAGCAGATCAAG 751
600 leuProSerArgLysSerGlnleuValleuValleuValleuSerAsp 616
752 CTACCCCGCTCTGTAAGAGCGAATCTGTAGTGAAGTAACTAGATTA 801
616 slleuProGlnGluMetGlyleuLeuGlnGlnGlnSerGlyAspLys 633
802 GCTCCAGAGAGATGGGCTCTGCGAGGCGACACGCGTGACAGAGG 851
633 laProGlyAspGlnGlyLysTyrrlleuAspleuArgHis 646
852 CTCGGGAGACAGGGGAGAAATACATCATCTGAGGCAT 892

seq_name: gb_est1:AL545925
seq_documentation_block:
LOCUS AL545925 879 bp mRNA linear EST 16-FEB-2001
DEFINITION AL545925 lYI_NFL006.PL2 Homo sapiens cDNA clone CSDD1023YF23 5
prime, mRNA sequence.
ACCESSION AL545925
VERSION AL545925.1 GI:12878563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD1023YF23"
/clone_1lb="lYI_NFL006.PL2"
/issue_type="Placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-Oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 226 a 216 c 276 g 157 t 4 others
ORIGIN
```

```
alignment_scores:
Quality: 1480.50 Length: 293
Ratio: 5.158 Gaps: 2
Percent Similarity: 97.952 Percent Identity: 97.611
```

```
alignment_block:
US-09-653-961-2 x AL545925
```

Align seg 1/1 to: AL545925 from: 1 to: 879

```
153 LysGlnProGlnGluValAlaThrCysValGlyArgAsnGlyTyrrProI 169
2 MAGAGACCTGAGAGAGTGGCTGCTACTGTGTAGGAGAGAACGGGTACCCAT 51
169 eProGlnVallleTyrrLysAsnGlyArgProLeuLysGlnGluLysA 186
52 TCCTCAAGTCATCTGTGACCAAGTAATGGCGGCTCTGAAAGAGGAGAA 101
186 snArgValHislleGlnSerSerGlnThrValGlnSerSerGlyLeuTy 202
102 ACCGGGTCCACATTCATGCTGCTCCAGACTGTGAGTGCAGTGTGTGTAC 151
203 ThrLeuGlnSerlleLeuLysAlaGlnleuValleuLysGlnAspLys 219
152 ACCTTCAGAGATATCTGAAGGCACACAGCTGTAAAGAAAGCAAGATGC 201
219 eGlnPheTyrrCysGluLeuAsnTyrrArgLeuProSerGlyAsnHis 236
202 CCAGTTTACTGTGTGACTACACTACCGGCTGCCAGTGGAAACCATGA 251
236 YsgLysSerArgGluValThrValProValPheTyrrProThrGlnLys 252
252 AGAGATCCAGGAGATCACCGTCCCTGTCTTCTACCCGACAAATAAGTG 301
253 TrpLeuGlnValGluProValGlyMetLeuLysGlnGlyAspArgValG 269
302 TGCTGTGAAGTGGAGCCCGTGGAAATCTGAAGAGAGGGGAGCGGTGCA 351
269 ulleArgCysleuAlaAspGlyAsnPropProPheHisPheSerlle 286
352 AATCAGGTGTGTGGCTGATGGCAACCTCCACACACTTCAGCATCAGCA 401
286 YsgLysAsnProSerThrArgGlnAlaGlnGlnGlnThrThrAsnAs 302
402 AGCAGAACCCAGCACAGGAGGCAAGAGAAAGCAACACACACACAC 451
303 GlyValleuValleuGluProAlaArgLysGlnHisSerGlyArgTy 319
452 GGGGTCTGTGTGTGTGACCTGCGCCGAGAAAGAACACAGTGGGCTAT 501
319 uCysGlnGlyLeuAspLeuAspThrMetlleSerlleuLeuSerGlu 336
502 ATGTACAGGCGCTGGACTTGGACACCATGATATGCTGCTGATGAA 551
336 InGlnleuValleuValAsnTyrrValSerAspValArgValSerPro 352
552 AGGAATCTACTGTGAATATGTGTGACGTCCGAGTGAAGTCCCGCAC 601
353 ProGlnArgGlnGlySerSerleuThrleuThrCysGlnAlaGln 369
602 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651
369 rSerGlnAspLeuGlnPheGlnTrpLeuArgGlnGlnThrGlyGlnVal 386
652 TACCCAGGAGCTCGAGTTCAGTGGCTGCTGAGARARARARAGAGG 701
386 euGlnArgGlyProValleuGlnleuHisAspLeuLysArgGlnAla 402
702 TGGAAGGGGGGCTGTGCTTCAAGTTCATGATGAAAGGAGGAGAGAG 751
403 GlyGlyTyrrArgCysValAlaSerValProSerlleProGlyLeu 419
752 GCGGCGTATCGCTGCGTGGCGTGTGCCAGCATACCGGCGCTGAA 801
419 gThrGlnleuValleuValleuValleuValleuValleuValleu 435
802 CACACAGCTGTGACAGCTGCA...TTTTTGGCCCTTGATGATGCATTC 848
436 LysGlnArgLysValThrValLysGln 444
849 AAGGAAAGAGAGGTGTGTGTAATAAAAA 875
```

seq_name: gb_est1:AL548612
seq_documentation_block:
LOCUS AL548612 L1.NFL006.PL2 Homo sapiens cDNA clone CS001036YB01.5
DEFINITION AL548612 L1.NFL006.PL2 Homo sapiens cDNA clone CS001036YB01.5
prims: mRNA sequence.
ACCESSION AL548612
VERSION AL548612.1 GI:12883791
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 868)
AUTHORS L.L.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS001036YB01"
/clone_lib="L1.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 188 a 249 c 269 g 160 t 2 others
ORIGIN

Alignment_scores:
Quality: 1426.50 Length: 288
Ratio: 5.023 Gaps: 2
Percent Similarity: 98.611 Percent Identity: 98.611

Alignment_block:
US-09-653-961-2 x AL548612 ..
Align seg 1/1 to: AL548612 from: 1 to: 868

```
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaIaCysCysC 17
|||||
12 ATGGGGCTTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61
|||||
17 sGcysProArgValAlaIaGlyValProGlyGluAlaGluInProAlaProG 34
|||||
62 CTGTCTCTCCGCTGCGGGGTGCGCCGAGAGCTGAGCAAGCTCGGCTG 111
|||||
34 luleuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
112 AGCTGTGGAGGTGGAGTGGAGCGACGAGCCCTCTCAATGTGGGCTC 161
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisI 67
|||||
162 TCCCACTCCCAAGGCCAACCTCAGCCATGTCGACTGTTTCTGCCACA 211
|||||
67 sGluysArgThrLeuIlePheArgValArgGlnGlyGlnGlyGlnSerG 84
|||||
212 GGAGAAAGCGAGCGTCACTCTCGTGTGGCCAGGCGCAGGCGCAGAGCG 261
```

```
84 lUpProGlyGlnTrpGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
262 AACCTGGGAGTACAGACGAGCGCTCAGCTCCAGACAGAGGGCTACT 311
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuGly 117
|||||
312 CTGGCCCTGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 361
|||||
117 ngIlyArgProArgSerGlnGlnTrpArgIleGlnLeuArgValTrp 134
|||||
362 GGGCAAGCGCCCTCGCTCCAGAGTACCGCATCCAGCTCCGCTCCACA 411
|||||
134 ysaIaProGlu.GluProAsnIleGlnValaAsnProLeuGlyIlePro 150
|||||
412 AACCTCGGAGGSCA...ANCATCAGATCAACCCCTGGGAGCATCCCTG 458
|||||
150 lAsnSerLysGluProGlnGluValAlaThrCysValGluArgAsnGly 167
|||||
459 GAACAGTAAAGAGCCTGAGAGGTCCGTACCTGTAGAGAGAACGGGT 508
|||||
167 yrrProIleProGlnValIleTrpTrpLysAsnGlyArgProLeuGly 183
|||||
509 ACCCATCTCTCAAGTCACTGTGGTACAGAAATGGCCGCTCTGAAGAG 558
|||||
184 GluLysAsnArgValHisIleGlnSerSerGlnTrpValGluSerSer 200
|||||
559 GAGAAAGACCGGCTCCACATTCATGCTCCAGACTGTGAGTGAAGTGA 608
|||||
200 yLeuTrpThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAs 217
|||||
609 TTGTGTACACTTCGACAGATATTCGAAGCAGACAGCTGTAAAGAAACA 658
|||||
217 ysaSplAglInPheTrpCysGluLeuAsnTrpArgLeuProSerGlyAs 233
|||||
659 AAGATCCAGCTTACGTAGAGTCACTACCGGCTCCAGTGGGAA. 707
|||||
234 HisMetLysGlnSerArgGluValThrValProValPheTrpProTr 250
|||||
708 CACATGAAGAGTCCAGAGAGTACCGCTCTGTTCATCCAGAGAGA 757
|||||
250 ulysValTrpLeuGluValGluProValGlyMetLeuLysGluLysAs 267
|||||
758 AAAAGTGTGCTGAGAGTGGAGCCCTGGGAATGCTGAAGAGAGAGGAG 806
|||||
267 rGValGluIleArgCysLeuAlaAspGlyAsnProProHisPheSer 283
|||||
807 C.GTGGAAATCAGGTGTTGGCTGATGCAACCTCCACCACTTACG 856
|||||
284 lIleSerLysGln 287
|||||
857 ATCAGCAAGCAG 868
```

seq_name: gb_est1:AL569913
seq_documentation_block:
LOCUS AL569913 L1.NFL006.PL2 Homo sapiens cDNA clone CS001001YH03.5
DEFINITION AL569913 L1.NFL006.PL2 Homo sapiens cDNA clone CS001001YH03.5
prims: mRNA sequence.
ACCESSION AL569913
VERSION AL569913.1 GI:12925725
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1019)
AUTHORS L.L.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
SOURCE

Location/Qualifiers
1. 1019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS01016YD21"
/clone_lib="PLT.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 243 a 292 c 291 g 189 t 4 others

ORIGIN

alignment_scores:
Quality: 1424.00 Length: 281
Ratio: 5.122 Gaps: 0
Percent Similarity: 98.932 Percent Identity: 98.577

alignment_block:
US-09-653-961-2 x AL569913 ..

Align seg 1/1 to: AL569913 from: 1 to: 1019

```

366 lu1aaguserSerGlnaSpluEglnpHeGlnTrrpleuArgGluGln 382
|||||:::|||||
44 GAATTCCGGGATGCCAGACCTGAGTCCAGTGGCTGAGAGAGAGAC 93
382 rGlyGlnValleuGlnArgGlyProValleuGlnleuHisAspLeuLys 399
|||||:::|||||
94 AGCCGAGTCTGGAAAGGGGCTGTGCTCAGTTCAGTACACTGAAC 143
399 rGluAlaGlyGlyGlyTyrArgGlyValAlaSerValProSerLlePro 415
|||||:::|||||
144 GGGAGGCGAGAGGGGCTATCGTGGGCTGTGCCAGCATATCCC 193
416 GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTr 432
|||||:::|||||
194 GGCTGAACCCGACACACTGTCACAGTGGCAATTTTGGCCCCCTTG 243
432 pheAlaPheLysGluArgLysValTrpValLysGluAsnMetValLeu 449
|||||:::|||||
244 GATGGCATTCAGAGAGAGAGAGTGTGGTCAAGAAATATGTTGA 293
449 snLeuSerGlyGluAlaSerGlyHisProArgProThrLleSerTrpAsn 465
|||||:::|||||
294 ATCTGCTGTGAGAGCTCAGGGACCCCGGCCCATCTCTCTGAGAC 343
466 ValAsnGlyThrAlaSerGlnAspGlnAspProGlnArgValLeuSe 482
|||||:::|||||
344 GTCAAGCGCAGCAAGTGAACAAGATCCACAGCGAGTCTCTGAG 393
482 rThrLeuAsnValLeuValThrProGluLeuLeuGlnThrGlyValGluC 499
|||||:::|||||
394 CACCTGAATGTCCTCGTGAACCCCGAGCTGTGGAGACAGGTGTGAT 443
499 ysrThrAlaSerAsnAspLeuGlyLysAsnThrSerLleLeuPheLeuGlu 515
|||||:::|||||
444 GCAGCGCTCCAGAGACTGGGCAAAACACACGATCTCTCTCTGAG 493
516 LeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyLe 532
|||||:::|||||
494 CTGTCAATTTAAACACCTCACACAGATCCACACACACCTGGCCT 543
532 userThrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrG 549
|||||:::|||||

```

544 CAGCACTTCCACTGCCAGTCTCTATACAGAGCCAAACAGCACTCCACAG 593

549 luArgLysLeuProGluLeuProGluSerArgGlyValValIleValAlaVal 565

594 AGAGAAAGCTCCGGAGCCGAGAGCCGGGGCGTGTGATCGTGGCTGTG 643

566 lleValCysIleLeuValLeuAlaValleuGlyAlaValleuTyrPhe 582

644 ATTTGTGATCTCTGCTGCTGGCTGTGGCTGTGGCTGTCTCTATTTCT 693

582 utyrTrpLysGlyLysLeuProCysArgArgSerGlyLysGlnGluIle 599

694 CTTATAGAGGCGCAGCTGCCGTGAGCGCTCAGGAGGAGCAGAGATCA 743

599 hrLeuProProSerArgLysSerGluLeuValAlaGluValLysSerAsp 615

744 CGCTGCCCGCTCTGTAGAGCGAATCTTGATTTGAAGTTAATCATGAT 793

616 LysLeuProGluGluMetGlyLeuLeuGlnLysSerSerGlyAspLysAr 632

794 AACCTCCAGAGAGATGGGCTCTCTGAGGCGACAGCGGTGACAGAG 843

632 galAProGlyAspGlnGlyGluLysTrpLleAspLeuArgHis 646

844 GGCTCCGGGAGACAGGAGAGAAATCATGATCTGAGGCAAT 886

seq_name: gb_est1:AL548357

seq_documentation_block:
LOCUS AL548357 864 bp mRNA linear EST 16-FEB-2001
DEFINITION AL548357 LTI.NFL006.PL2 Homo sapiens cDNA clone CS01016YD21 5
prime, mRNA sequence.
ACCESSION AL548357
VERSION AL548357.1 GI:12883284
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 864)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
SOURCE

1. 864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS01016YD21"
/clone_lib="PLT.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 187 a 249 c 265 g 160 t 3 others

ORIGIN

alignment_scores:
Quality: 1356.50 Length: 279
Ratio: 4.951 Gaps: 3
Percent Similarity: 98.208 Percent Identity: 98.208

```
alignment_block:
```

US-09-653-961-2 X AL548357

Align seg 1/1 to: AL548357 from: 1 to: 864

1 MCTUYLEUPROARGLEUVALCYSLALAPHEULEUVALAALACYSCYCY 17
 4 ATGGGGCTTCCAGGCTGTGCTGGCCCTTCTTGGCTGGCCGCTGCTG 53
 17 GCYAPROARGVALAAGLYVALPROGLYGLAAGLUGLPROBLAPROG 34
 54 CTGTCTCTGGCTGGGGGTGTGGCCGGAGAGCTGACAGCCTTCCGCTG 103
 34 LULEUVALIGLVALIGLVALIGLYSERTHRALALEULEULYSCYGLYLEU 50
 104 AGCTGTGTGAGGTGGAAGTGGGACAGACAGCCCTTCTGAAGTGGGCTC 153
 51 SERGLNSERGLNLYAENLSENRHIVALSPPRPHESERVALHSLY 67
 154 TCCCAAGTCCCAAGGCACACTCAGCCATGTGACTGGTTTCTGTCCAA 203
 67 SGULYARGTHRILEULEPHARGVALARGGLNGLYGLNGLNSERG 84
 204 GGAGAAACGGGACGCTCATCTTCCGTGGTGGCCCAAGGGCCAGAGGCG 253
 84 LUPROGLYGLTYRGLUGLARGLEUSERLEUGLASPARGLYALAH 100
 254 AACCTGGGGAGTACGAGACAGGCTCAGCTTACCAGACAGAGAGGCTAC 303
 100 FLEUALALEUTHGLVALTHPRGLNASPLUARGILEPHELEUCYAG 117
 304 TCTGGCCCTGACTCAAGTACCCCCCAAGAGGCCATCTTCTGTGCC 353
 117 LUGLYLYSARGPROARGSERGLNGLYTARGILEGLNEUARGVALTYR 133
 354 AGGGCAAGGCGCCTCGGCTCCAGAGATACGCATCAGCTCGGCTTAC 403
 134 LYSALAPROGLUGLPROASNLGLVALASNPROLEUGLYLEPROV 150
 404 AAAGCTCGGAGGSCA..AACATCAAGTCAACCCCTGGGCAATCCCTG 450
 150 ALASNSERTYSLUPROGLUGLVALALATHCYSAVALIARGASNLY 166
 451 TCAACAGTAAGAGGCTGAGGAGTTCGCTACTGTTGAGGGAGAACGG 500
 167 TYRPROILEPROGLIVALIETHPRYTYRYSNGLYARGPROLEULYSGI 183
 501 TACCCATTTCCCAAGTACTGTGGTACAAGATGGCCGCTCTGAAGA 550
 183 UGLYLYSASNAARGVALHISILEGLNSERSEGLNTHYVALIGLUSERSERG 200
 551 GGAGAAAGAACGGGCGCCACATTCAGTGCATCCACAGCTGTGAGTGCAGT 600
 200 LYLEUTYRTHLEUGLNSERILEULEYSLAAGLNEUVALYSGULASP 216
 601 GTTGTACACCTTCCAGAGTATTTGTGAAGGACAGCTGTAAAGAAGAC 650
 217 LYSASPALAAGLNIETHYRYSGLULEASNPYTRARGLEUPROSERGLYAS 233
 651 AAACATGGCCCGATTACTGTAGACTCACTACACCGGCTGCCCATGTGGAA 699
 233 NHIMETLYSEGLUSERARGVALIHPVALPROVALPHERTYRPROTHR 250
 700 CACATGAAGAGATCCAGGGAAGTCAACCGCTGTGTTTCTTACCCAGAC 749
 250 LULYSVALTPLEUGLVALIGLUPROVALIGLYMETLEULYSGULIYASP 266
 750 AAAAAGTGTGGCTGGAAGTGGACCCGCTGGGAATGCTGAAGAGAGGAGC 799
 267 ARGVALIGLILEARGCYSLAUALAASPGLYASN 277
 800 GC.GTGGAAATCAGGCTTTGGCTGTATGGCAAC 831

seq_name: gb_est1:AU119585

```
seq_documentation_block:
```

LOCUS DEPTA

DEFIN

Appendix

ACCESS
FIRST

KEYWORDS

SOURCE

ORG

...

REFER

AUT

34

TIT

YOU
CANNOT

COMME

FEATURES

50

BASE C

ORIGIN

alignm

	Percent Similarity:	5.028	Ratio:	2
	Percent Identity:	97.786	Gap:	2
			Percent Identity:	96.679

Alignment block:
 09-653-961-2 x AU119585 ..

Alignment segment 1/1 to: AU119585 from: 1 to: 835

1	MetGlyLeuProArgLeuValCysAlaPheLeuAlaAlaCysCysC	17
19	ATGGGGCTCCACAGGCTGGCTCCGCCCTTCGCTCGCGCGCTGCTG	68
17	scsProArgValAlaGlyValProGlyValIleValInProAlaProG	34
69	CTGTCCTGGGCTGGGGGTGGTCCCGAAGGCTGAGCAGGCTCGGCTG	118
34	IleValAlaGluValAlaGluValGlySerThrAlaLeuLeuLeuCysGlyLeu	50
119	AGCGTGGTGGAGGTGAAGTGGGACAGACAGCCCTTCGTAACTGGCGGCTC	168
51	SerGlnSerGlnGlyLysLeuSerHisValAspTrpPheSerValHisI	67
169	TCCCAAGTCCCAAGGCAACCTCAGCCATGTCACAGCTTCTGTCACAA	218
67	scGlyLysArgThrLeuIlePheArgValArgGlnGlyGlnGlyLysSerG	84
219	GGAGAAAGCGGACGCTATCTTCCTGCTGGCCCAAGGCGCAAGGCGCAAGG	268

84 lnpProGlyGluTyrGluGlnArgLeuSerLeuGlnAspArgLysAlaThr 100
|||||
269 AACCTGGGAGTACGACGCGCTCAGCTCCAGACGAGGCGCTACT 318
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgLysPheLeuGly 117
|||||
319 CTGGCCCTGACTCAAGTACACCCCAAGAGAGAGCATCTCTGTGCCA 368
|||||
117 nGlyAspArgProArgSerGlnGluTyrArgLysGlnLeuArgValTyrL 134
|||||
369 GGGGAGGCGCCCTCGGTCCAGAGTACCGCATCCAGCTCCCGCTTACA 418
|||||
134 ysaAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
|||||
419 AACCTCCGGAGGAGCCAAACATCCAGGTACACCCCTGGGATCCCTGG 468
|||||
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTyr 167
|||||
469 AACAGTAAAGAGAGCTGAGAGGTGCTGCTGTTAGGAGGAGAGCGGTA 518
|||||
167 rProIleProGlnValIleTyrTyrLysAsnGlyArgProLeuLysGluG 184
|||||
519 CCCCATCTCTCAAGTATCTGTGACAGAAATGCGCGCTCTGAAAGAGG 568
|||||
184 lAlaAsnArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
|||||
569 AGAAGAACCGGGTCCACATTCAGTGTCCAGAGCTGTGAGCGAGTGT 618
|||||
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
|||||
619 TTGTACACCTTGCGAGTATTTCTGAAGGACACAGCTGTTAAAGAACAA 668
|||||
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnH 234
|||||
669 AGATGCCAGTTTACTGTGAGCTCACTACAGCGCTGCCANTGGGAGAC 718
|||||
234 isMetLysGluSerArgLysValThrValProValPheTyrProThrG 250
|||||
719 ACATGAAGGAATTCAGGGAGTCAACCGC...CTGTTCTACCCGACAG 765
|||||
250 ulysValAlaTrpLeuGluValGluProValGlyMetLeuLysGluGly 267
|||||
766 AAAAGTGTGCTGGAATG.GAACCGTGGGAATGCTGAAGGAAGGAGCC 814
|||||
267 rGlyAlaGluIle 270
|||||
815 GCGTGGAAATA 825
|||||
seq_name: gp_est1:AU119127
seq_documentation_block:
LOCUS AU119127 743 bp mRNA linear EST 19-OCT-2000
DEFINITION AU119127 HEMBA1 Homo sapiens cDNA clone HEMBA1005093 5', mRNA
sequence.
ACCESSION AU119127
VERSION AU119127.1 GI:10934362
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE
HRI human cDNA project
JOURNAL
Unpublished (2000)
COMMENT
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3851
Fax: 81-438-52-3952

Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute: cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. 743
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1005093"
/clone_11b="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="Vector: pMT18SFL3"
BASE COUNT 185 a 189 c 238 g 128 t 3 others
ORIGIN
alignment_scores:
Quality: 1244.00 Length: 245
Ratio: 5.205 Gaps: 1
Percent Similarity: 97.551 Percent Identity: 97.551
alignment_block:
US-09-653-961-2 x AU119127 ..
Align seg 1/1 to: AU119127 from: 1 to: 743
216 AsplysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerG 232
|||||
3 GACAAAGTGGCCAGATTCTGAGTGTGAGTCAACACCGCGCTGACCAATGG 52
|||||
222 yAsnHisMetLysGlnSerArgGluValThrValProValPheTyrPro 249
|||||
53 GACACCATGAAAGAGTCCAGGAAAGTACCGCTCCCTGTTTCTACCCGA 102
|||||
249 hGlyLysValAlaTrpLeuGluValGluProValGlyMetLeuLysGly 265
|||||
103 CAGAAAGAGTGGCTGAGTGAAGTGGAGCCCGTGGGAATGCTGAAGAGAGG 152
|||||
266 AspArgValGluIleArgCysLeuAlaAspGlyAsnProProProHisPh 282
|||||
153 GACCCGCGGAATTCAGGTGTTGGCTGATGCAACCTCCACCACTT 202
|||||
282 eSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluGluThr 299
|||||
203 CAGCATCAAGCAAGCAAGACCCAGACAGGAGGAGGAGCAAGAACAA 252
|||||
299 hAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSer 315
|||||
253 CCAACGACAAACGGGGTCTGCTGCTGAGAGCTGCCCGGAAAGAACACAGT 302
|||||
316 GlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 332
|||||
303 GGGCGCTATGAATGTCAGGCGCTGACCTTGACACCAATGATATGCTGCT 352
|||||
332 uSerGluProGlnGluLeuLeuValAsnTyrValSerAspValArgLys 349
|||||
353 GAGTAAACACAGCAAGACTACTGTGTAATATGTGTCTACGTCGAGATGA 402
|||||
349 eProAlaAlaProGluArgGlnGlnGlySerSerLeuThrCys 365
|||||
403 GTCGCGAGCCCTGAGACAGAGAAAGCAGACCTCACCTGACTGCTGT 452
|||||
366 GluAlaGluSerSerGlnAspLeuGluPheGlnThrPheArgGluGluThr 382
|||||
453 GAGGAGGAGAGTACGACGACCTCGAGTTCAGATGGCTGAGAGAAAGAC 502
|||||
382 rGlyGlnValLeuGlnArgGlyProValLeuGlnLeuHisAspLeuLys 399
|||||
503 AGCCAGAGTGTGGAAGAGGGGCGCTGCTCAGTGTGATGATACCTGAAC 552
|||||
399 rGlyAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415

```
|||||
553 GGGAGCGAGGCGGCTATGCTGCGCTGTGCCCCACCATACCC 602
416 GYLeuAenAArgThrGlnLeuValAsnValAlaIlePheGlyProProTr 432
603 GGCCTGAACCCACACACAGCTGTCACCTGCCCCCTTTTGGCCCCCTTG 652
432 pMetaIaPheLysGluArgLysValATrPValLysGluAsnMetValLeuA 449
653 GATGGCATTTCAAGAGAGAGAGAGTGTGGTGGAAGANATATGNGTTGA 702
449 snLeuSerCysGluAlaSer GlyHisProArg 459
703 ATCTGCTTGNGAAGCCTTAAGGACCCCGG 735

seq_name: gb_est1:AU126003

seq_documentation_block:
LOCUS AU126003 735 bp mRNA linear EST 23-OCT-2000
DEFINITION AU126003 NT2RM4 Homo sapiens cDNA clone NT2RM4002547 5', mRNA
sequence.
ACCESSION AU126003
VERSION AU126003.1 GI:10950719
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 735)
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
,Y., Sugano,S., Isogai,T.)
Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomise@hri.co.jp
HRI human cDNA project, 5'- & 3'-end one pass sequencing: Helix
Research Institute: cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4002547"
/clone_11b="NT2RM4"
/cell_type="teratocarcinoma"
/note="Vector: pMet18SFL3, mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 192 a 191 c 220 g 129 t 3 others
ORIGIN

alignment_scores:
Quality: 1238.00 Length: 244
Ratio: 5.116 Gaps: 0
Percent Similarity: 99.180 Percent Identity: 99.180

alignment_block:
US-09-653-961-2 x AU126003 ..

Align seq 1/1 to: AU126003 from: 1 to: 735
131 ArgValTyrLysAlaProGluGluProAsnIleGlnValAsnProLeuG1 147
|||||
```

```
3 CCGCTCTACAAAGCTCCGGAGAGCCAAACATCCAGTCAACCCCTGGG 52
147 YIleProValAsnSerLysGluProGluGluValAlaIleThrCysValGly 164
53 CATCCCTGTGAACAGTAAGAGCTGAGAGGTGCGTACTGCTGTAAGGSA 102
164 rGaSnGlyTrProIleProGlnValIleTrPyrLysAsnGlyArgPro 180
103 GGAACGGGTACCCCATTTCCCAAGTCATCTGTGACAGAAGATGGCGGCT 152
181 LeuLysGluLulysAsnArgValHisIleGlnSerSerGlnThrValG1 197
153 CTGAAGAGAGAGAAGAACCGGGTCCACATTCAGTGTGCTCCACAGACTGGA 202
197 userSerGlyLeuTrThrLeuGlnSerIleLeuLysAlaGlnLeuValL 214
203 GTTCAGTGTGTTGTACACCTTGACAGATTTCTGAAGGACGACGTGGTTA 252
214 ysgLusPlysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuPro 230
253 AAGAAGACAAAGATGCCAGTTTACTGTGAGCTCAACTACGGCGCTGCC 302
231 SerGlyAsnHisMetLysGluSerArgGluValThrValProValPheTy 247
303 AGTGGAGAACCATGATGAAGAGATCCACAGGAATCACCGCTCCGTTTCTA 352
247 rProThrGluLysValTrPLeuGlnValGluProValGlyMetLeuLysG 264
353 CCGGACAGAAAAAGTGGCTGGTGAAGTGAAGCCCGTGGGAATGCTGAAGG 402
264 LulysAspArgValGluIleArgCysLeuAlaAspGlyAsnProProPro 280
403 AAGGGGACCGCGGCAATCAGGTGTTGGCTGATGGCAACCTCCACCA 452
281 HisPheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGlu 297
453 CACTTGACATCAGCAAGCAGAACCCACACAGGAGAGGAGAGAA 502
297 uThrThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluH 314
503 GACCAACCAAGCAACGAGGGGCTCTGCTGTGAGCGCTGCCGGAAGAAC 552
314 isSerGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSer 330
553 ACAGTGGCGCTATGATGATGTCAGGGCTGGAAGTGAACCATGATATCG 602
331 LeuLeuSerGluProGlnGluLeuValAsnTyrValSerAspValAr 347
603 CTGCTGATGAACCAACGAGAACTACTGGTGAATGTGTCTGACGCTCG 652
347 gValSerProAlaAlaProGluArgGlnGluGlySerSerLeuThrLeuT 364
653 AGTGAATCCCGCAGCGCCCTGAGAGACAGGAAGACGACGCGC. ACCCTGA 701
364 hCysGluAlaGluSerSerGlnAspLeuGlu 374
702 CCTGTGANCANANAGTACCCAGACCTCGAG 733

seq_name: gb_est1:AL550956

seq_documentation_block:
LOCUS AL550956 923 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550956 LIT1_NFL006_Pu2 Homo sapiens cDNA clone CS0D1065YU10 5
prime, mRNA sequence.
ACCESSION AL550956
VERSION AL550956.1 GI:12888440
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 923)
L1,W.B., Gruber,C., Jessee,J. and Polayes,D.
```


TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Source 1..923
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSOD1065XJ10"
/clone_lib="LIT_NFL006_PL2"
/lssue_type="Placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 202 a 212 c 268 g 192 t 49 others
ORIGIN

alignment_scores:

Quality: 1221.00 Length: 304
Ratio: 4.556 Gaps: 1
Percent Similarity: 88.158 Percent Identity: 84.868

alignment_block:
US-09-653-961-2 x AL550956 ..

Align seg 1/1 to: AL550956 from: 1 to: 923

```
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaLacCysCysC 17
|||||
1 ATGGGGCTTCCAGGCTGGTCTCGCCTTCTGCGCCGCGCTGCTG 50
17 scysProArgValAlaGlyValProGlyValAlaGlyValProAlaProG 34
|||||
51 CTGTCTCTGCGCTGCGGCTGCTGCGGAGGCTGAGCACCCTGCTG 100
34 IuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
101 AGCTGCTGAGGTGAGAGTGGGACACACACCTTCTGAAAGTCCGCTC 150
51 SerGlnSerGlnGlnValAsnLeuSerHisValAlaSerTrpPheSerValHis 67
|||||
151 TCCCACTCCCAAGGCAACCTCAGCAGCTGACTGTTTCTGTCACAA 200
67 sGlyLysArgThrLeuIlePheArgValArgGlnGlnGlnGlnSerG 84
|||||
201 GGGGAAGCGGAGCGCTCATCTCCGTGTGGCCAGGGCCAGGCGCAAGCG 250
84 IuProGlyLysTrpGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
251 AACCTGGGAGTAGACGACGCGCTCAGCTCCAGANANAGGGCTACT 300
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCysG 117
|||||
301 CTGGCCCTGANTCAAGTACCCCAAGATGACGATATTTTTCGCGCA 350
117 nGlyLysArgProArgSerGlnGlnLysArgIleGlnLeuArgValTyr 134
|||||
351 GGGCAAGCGGCTCTGGTTCAGGAAKTAAGCAATYAGTTCGGCTTAYA 400
134 ysaLapProGlnGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
|||||
401 AAKYTCGGGAGATTA. AAYATTYAGTTAACTCCCTGGGCAATYTTGTT 449
```

```
151 AsnSerLysGluProGlnGluValAlaThrCysValGlyArgAsnGlyTyr 167
|||||
450 AATAKTAAGGAGCTTGAAGAGTGGCTACTGTTGTAGGAGGAATGCTGA 499
167 rProIleProGlnValIleTrpTyrLysAsnGlyArgProLeuLysGluG 184
|||||
500 TCCCATYTYTYAATTATTTTGTACAGAAATGTTGTTCTGAMGATG 549
184 IuysAsnArgValHisIleGlnSerSerGlnTrpValGluSerSerGly 200
|||||
550 AGAGMWTTCGGTCTATATTAGTGTCCAGACTGTGAGTGAATGATGT 599
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
|||||
600 TTGTACACCTTGYAGAGTATTTCAAGCAGCAGTGTTAAGAAAGCAA 649
217 sasPalaGlnPheTyrCysLysLeuAsnTyrArgLeuProSerGlyAsn 234
|||||
650 AGATGCCCAATTTTACTGTAGCTCACTACCGGCTGCCAGTGGGAA. C 698
234 IsmLysGlnSerArgGluValThrValProValPheTyrProThrGlu 250
|||||
699 ACATGAAGAGTCCAGGAAAGTACCGTCCCTGTTTCTACCGAATGAA 748
251 LysValTrpLeu. GluValGluProValGlyMetLeuLysGluLysPa 267
|||||
749 AAAGTGTGGTGGGAATGKAGYCCGCGGGAATGCTTAAGAAAGGAGC 797
267 rGValGluIleArgCysLeuAlaAspGlyAsnProProHisPheSer 283
|||||
798 C.GTGGAAATCAGGTGTTGCTGATGCAACCTCCACACACTTCAGC 847
284 IleSerLysGlnAsnProSerThrArgGluAlaGlnGluLysTrpTrp 300
|||||
848 ATCAGCAGACAGAAACCCAGCACCGAGGAGCAGAGAAAGCAACAC 897
300 nasPasnGly 303
|||||
898 GACACAGCGG 907
seq_name: gb_est1:AU130835
seq_documentation_block:
LOCUS AU130835 735 bp mRNA linear EST 24-OCT-2000
DEFINITION AU130835 NT2RP3 Homo sapiens cDNA clone NT2RP3001526 5', mRNA
ACCESSION AU130835
VERSION AU130835.1 GI:10991189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 735)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genom@hri.co.jp
HRI human cDNA project; 5'- & 3'-end pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
location/Qualifiers
1..735
/organism="Homo sapiens"
```

/db_xref="taxon:9606"
/clone="NT2RP3001526"
/clone_id="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 155 a 221 c 221 g 135 t 3 others
ORIGIN

alignment_scores:
Quality: 1193.00 Length: 232
Ratio: 5.210 Gaps: 0
Percent Similarity: 98.707 Percent Identity: 98.276

alignment_block:
US-09-653-961-2 x AU130835 ..

Align seg 1/1 to: AU130835 from: 1 to: 735

```
1 MetGlyLeuProArgLeuValGlyAlaPheLeuAlaIaCysCysC 17
|||||
37 ATGGGGCTCCAGAGCTGCTGCGCTTCTGCTGCGCGCTGCTG 86
|||||
17 SCYPRoArgValAlaGlyValProGlyGluAlaGluInProAla 34
|||||
87 CTGTCTCGGCTCGCGGCTGCGCGAGAGGCTGAGCACTCGCGCT 136
|||||
34 LuLeuValGluValGlySerThraLeuLeuLeuLeuLeu 50
|||||
137 AGCTGGTGGAGTGGAAGTGGGAGCACACCTTCGAAAGTGGGCTC 186
|||||
51 SerGInSerGInGlyAsnLeuSerHisValAspTrpPheSerValHis 67
|||||
187 TCCAGTCCCAAGGCAACCTCAGCATGTGCTGCTTCTGCTCCACAA 236
|||||
67 sGluYsArgThrLeuIlePheArgValArgGInGInGInSerG 84
|||||
237 GGAGAACGGAGCGCTATCTTCGTGCGCCAGGCGGAGGCGAGGCG 286
|||||
84 LuProGlyGluTrpGluInArgLeuSerLeuGInAspArgGlyAlaThr 100
|||||
287 AACCTGGGGAGTACGACAGCGGCTCAGCTCCAGGACAGAGGCGCTACT 336
|||||
101 LeuAlaLeuThrGlnValThrProGInAspGluArgIlePheLeu 117
|||||
337 CTGGCCCTGACTCAAGTCAACCCCAAGAGAGCGCATCTTCTGTGCCA 386
|||||
117 nglyYsArgProArgSerGInGluTrpArgIleGInLeuArgValTrp 134
|||||
387 GGGCAAGCGGCGTGGTCCAGAGTACCGCATCAGCTCCGCGTACAA 436
|||||
134 ysaIaPProGluInProAsnIleGInValAsnProLeuGlyIleProVal 150
|||||
437 AAGCTCCGGAGAGCAAAACATCCAGGTCAACCCCTGGGCAATCCCTGTG 486
|||||
151 AsnSerYsGluProGluGluValAlaThrCysValGlyArgAsnGly 167
|||||
487 AACAGTAAGAGGCTGAGAGGTGCTACTGTGTAGGAGGAAGGAGTA 536
|||||
537 CCCCATTCCTCAAGTCACTGTGTACAAAGAAATGGCGGCTCTGAAAGAGG 586
|||||
167 rProIleProGInValIleTrpYrYsAsnGlyArgProLeuYsGlu 184
|||||
184 LuYsAsnArgValHisIleGInSerSerGInThrValGluSerSerGly 200
|||||
587 AGAAGAACCAGGCTCAGATTCACTGCTCCAGACTGTGAGTCAAGTGGG 636
|||||
201 LeuTrpThrLeuGInSerIleLeuYsAlaGInLeuValYsGluAsp 217
|||||
637 TTGTACACCTTGCAGAGTATTCTGAAGGCACTGGTTTAAGAGACAA 686
```

217 sasAlaGInPheTrpYrYsGluLeuAsnTrpArgLeuProSerGly 232
|||||
687 ACATGCCCAATTTTACTGAGACTCACTACCGGTTGCANTGGGA 732

seq_name: gb_est1:AL546898

seq_documentation_block:

LOCUS AL546898 773 bp mRNA linear EST 16-FEB-2001
DEFINITION AL546898 LTL_NFL006.PL2 Homo sapiens cDNA clone CS0D1026YB05 5
prime, mRNA sequence.

ACCESSION AL546898
VERSION AL546898.1 GI:12880463

KEYWORDS
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 773)
AUTHORS 'L.W.B., Gruber,C., Jessee,V. and Polyes,D.

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

FEATURES

source

1..773
/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0D1026YB05"

/clone_id="LTL_NFL006.PL2"

/issue_type="placenta"

/note="Vector: PCWSPORT 6; Site-1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWSPORT 6

vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 166 a 227 c 205 g 163 t 12 others
ORIGIN

alignment_scores:
Quality: 1159.50 Length: 255
Ratio: 4.872 Gaps: 2
Percent Similarity: 93.333 Percent Identity: 90.588

alignment_block:
US-09-653-961-2 x AL546898 ..

Align seg 1/1 to: AL546898 from: 1 to: 773

```
1 MetGlyLeuProArgLeuValGlyAlaPheLeuAlaIaCysCysC 17
|||||
11 ATGGKCTTCCAGAGCTGCTTCTTCCTTCTGCTGCGCGCTGCTG 60
|||||
17 SCYPRoArgValAlaGlyValProGlyGluAlaGluInProAla 34
|||||
61 CTGTCTCGGCTCGCGGCTGCTCCGGAAAHGCTAGAGCGCTGCGCTT 110
|||||
34 LuLeuValGluValGlySerThraLeuLeuLeuLeuLeu 50
|||||
111 ATCTGTTGAGAGTGGAAATTKGGCAGACAGCCCTTTAAGTKGGCCTC 160
|||||
51 SerGInSerGInGlyAsnLeuSerHisValAspTrpPheSerValHis 67
|||||
161 TCCAGTCCCAAGGCAACCTCAGCATGTGCTGCTTCTTCTTCACAAA 210
|||||
67 sGluYsArgThrLeuIlePheArgValArgGInGInGInSerG 84
```

```

|||||
211 GGAGAGAGCGAGCTCATCTCCGTGTGCGCCAGGGCCAGGGCCAGAGCG 260
84 lnpProGlylufyrglnArgleuSerleuGlnAspArgGlyAlaThr 100
|||||
261 AACCTGGGGAGTACGAGCAGCGGCTCAGCTCCAGCAAAAGGGCTACT 310
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuGlySGI 117
|||||
311 CTGGCCCTGACTCAAGTCAACCCCAAGACGAGCGATCTTCTTGCGCA 360
117 nGlylAspArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrL 134
|||||
361 GGCGAAGCTCCCTCGTCCAGGATTAACGCAATCCAGCTCCGCGCTACA 410
134 yAlaProGluGluPro.AsnIleGlnValAsnProLeuGlyIleProVa 150
|||||
411 AATCTCCG...GAGATTAACATCCAGTTCAACCCCTTGCGATCCCTTT 457
150 lAsnSerlYsgluProGluGluValAlaThrCysValGlyArgAsnGlyT 167
|||||
458 TTAACGTAAGAGAGCTTATGAGAGTCCCTTTTAAAGAGAGAGAGGCT 507
167 YrProIleProGlnValIleTyrTyrLysAsnGlyArgProLeuYsglu 183
|||||
508 ACCCATTCCTCAAAATCATCTGTACAAAGATGGCGGCGCTTAAAGAG 557
184 GluYsAsnArgValHisIleGlnSerGlnThrValGluSerSerI 200
|||||
558 GAGAGAACCCGGTCCACATTCAGTCCGACACTGTGGAGTCCAGTGG 607
200 YleuTyrThrleuGlnSerIleLeuLysAlaGlnLeuValYsgluAspL 217
|||||
608 TTTGTACACCTTCGACAGATTCCTGAGGACACAGCTGGTTAAAGAGACA 657
217 ySAspAlaGlnPheTyrCysGlnLeuAsnTyrArgleuProSerGlyAsn 233
|||||
658 AAGATGCCAGTTTACTGTGAGTCAACTACCGCGCTGCCAGTGGAA. 706
234 HisMetlYsgluSerArgGluValThrValProValPheTyrProThrL 250
|||||
707 CACATGAAGAGAGTCCAGGAGTCAACCGCTCTTTCTTACCCGACAGA 756
250 uLysValTyrLeu 254
|||||
757 AAAAGTGTGCTG 769

```

seq_name: qb_est2:BM083718

seq_documentation_block:

LOCUS BM083718 706 bp mRNA linear EST 16-NOV-2001
 DEFINITION ImageC2_2001/smk328bdf41.x1 NIH_MGC_16 Homo sapiens cDNA clone
 IMAGE:4634015 5', mRNA sequence.

ACCESSION BM083718

VERSION BM083718.1 GI:16951349

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Kale,P.I., Harsch,T.J., Folta,P.A., Nelson,D.O., Sanders,C.G. and
 Prange,C.K.

TITLE The I.M.A.G.E. Consortium quality control effort: clone
 resequencing for verification

JOURNAL Unpublished (2001)

COMMENT Other_ESTs: BG470560

Contact: Prange CK

The I.M.A.G.E. Consortium
 Lawrence Livermore National Laboratory
 Livermore, CA, USA
 Email: help@image.llnl.gov

This read has been verified (found to hit its original self in the

correct orientation), as part of the I.M.A.G.E. Consortium quality control effort. High quality sequence is defined as having 100 or more base pairs with a phred quality value of 20 or greater, where a sliding window of 4 base pairs with a phred quality value of 15 or greater marks the beginning and end of the sequence. For information on obtaining this clone, please contact info@image.llnl.gov.
 Plate: LICM1389 row: a column: 24
 Seq primer: -21m13
 High quality sequence stop: 706.

FEATURES

source

Location/Qualifiers

1..706

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4634015"

/clone_lib="NIH_MGC_16"

/tissue_type="retinoblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pOTF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT

143 a 214 c 217 g 131 t 1 others

ORIGIN

alignment_scores:

Quality: 1140.00

Ratio: 5.067

Percent Similarity: 97.826

Length: 230

Gaps: 2

Percent Identity: 96.522

alignment_block:

US-09-653-961-2 x BM083718

Align seg 1/1 to: BM083718 from: 1 to: 706

```

1 MetGlyLeuProArgleuValCysAlaPheLeuAlaLacCysCysC 17
|||||
21 ATGGGGCTTCCAGAGCTGTGCGCTCTTGTGCGCCGCGTGTG 70
17 scYsProArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
|||||
71 CTGTCTCGCGTCCGCGGTGCGCGAGAGGCTGAGACCTGCGCTG 120
34 lueValGluValGluValGlySerThrAlaLeuLeuYsglyLeu 50
|||||
121 AGCTGTGGAGGTGGAGAGTGGGCGACAGCCCTTGTGAAGTCCGGCTC 170
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTyrPheSerValHisIly 67
|||||
171 TCCAGTCCCAAGGCAACCTCAAGCCATGTGCACTGGTTTCTGTCCAAA 220
67 sGluYsArgTThrLeuIlePheArgValArgGlnGlyGlnGlyGlnSerG 84
|||||
221 GGAGAGCGGAGCGCTCATCTTCGTGTGCGCCAGGCGCGGCGCAAGGG 270
84 lnpProGlylufyrglnArgleuSerleuGlnAspArgGlyAlaThr 100
|||||
321 CTGGCCCTGACTCAAGTCAACCCCAAGACGAGCGCATCTTCTTGCGCA 370
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuGlySGI 117
|||||
117 nGlylAspArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrL 134
|||||
371 GGCGAAGCTCCCTCGTCCAGGATTAACGCAATCCAGCTCCGCGCTACA 420
134 yAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150

```

```
|||||
421 AACCTCCGAGAGACCAACATCCAGSTCAACCCCTGGGCATCCCTGTG 470
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArg.AsnGlyT 167
471 AACAGTAAGAGAGCTGAGAGGTCGTACCTGTGTAGGAGGGAACGGGT 520
167 YrProIleProGluValIleTyrLysAsnGlyArgProLeuLysGlu 183
521 ACCCATTTCTCTCAAGTCATCTGGTACAAAGAATGGCCGCCCTGTGAAGAG 570
184 GluLysAsnArgValHisIleGlnSerSerGlnThrValGlu.SerSerG 200
571 GAGAAAGAACCGGCTCCACATTCAGTCCGCAAACTGTGGAGTNCAGGTG 620
200 LysLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 216
621 GTTGTGACACCTTGACAGATATTCTGAAGGCACAGCTGTGTTAAAGAAAGA 670
217 LysAspAlaGlnPheTyrCysGluLeuAsnTyrArg 228
671 CAAGATGCCAGTTTACTGTGTGAGCTCACTACCGG 706
```